

THE CORRELATION BETWEEN
TEACHER CLARITY OF COMMUNICATION
AND STUDENT ACHIEVEMENT GAIN:
A META-ANALYSIS

By

FRANK FENDICK

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA 'UNIVERSITY OF FLORIDA LIBRARIES

1990

ACKNOWLEDGMENTS

It is my pleasure to acknowledge my debt for the guidance I have received over many years from my chairman, Dr. James Algina, and from the members of my committee, Dr. Wilson Guertin, Dr. Patricia Ashton, and Dr. Robert Ziller. Please accept my sincere thanks.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGMENTS	ii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
ABSTRACT.....	viii
CHAPTERS	
I INTRODUCTION.....	1
Teacher Clarity of Communication	1
Clarity of Speech	1
Clarity of Organization	1
Clarity of Explanation	2
Clarity of Examples and Guided Practice	3
Clarity of Assessment of Student Learning	3
Summary of the Dimensions of Teacher Clarity ...	3
The Problem Studied in This Dissertation	4
The Rationale for the Study	4
Practical Significance and Objectives of the Study	5
Theoretical Significance of the Study	7
Limitation of this Dissertation	9
Definitions	10
Definition of Teacher Clarity	10
Definition of Student Achievement Gain	11
Outline of the Dissertation	13
II REVIEW OF THE LITERATURE	14
Teacher Clarity Literature	14
The Dimensions of Teacher Clarity	14
Clarity of Speech	20
Clarity of Organization	22
Clarity of Explanation	23
Examples and Guided Practice	24
Assessment of Student Learning	25
Analysis Literature	25
General Problems of Meta-analysis	25
Criteria for Evaluating Meta-analyses	28
Criticisms of Meta-analyses in Education	29

	<u>Page</u>
III REVIEW OF METHODOLOGY	31
Objective	31
Uncertainty and Variance	32
Summary of Procedure	32
Defining the Problem	33
Finding the Studies	35
Describing, Classifying, and Coding Research Studies	36
The Dimensions of Teacher Clarity	36
Description and Classification	39
Coding Nominal Values	41
Techniques of Analysis	42
Looking at the Data	42
Problems in Accumulating the Effect Sizes From Each Study	43
Glass, McGaw, and Smith (1981)	46
Hunter, Schmidt, and Jackson (1982)	44
Rationale for the Hunter and Schmidt Method ...	45
Hedges and Olkin (1985)	48
Hedges (1988)	50
Summary of the Analyses Used in This Study	50
IV RESULTS AND ANALYSES	53
Results	53
Frequency Distribution	54
Removal of Outliers	54
Reliability of Dimensions of Teacher Clarity ...	58
Characteristics of the Reduced Data Set	58
Relationships Between Characteristics	61
Glass Analysis	64
Treating Each Effect Size as Independent	64
Using Tukey's Jackknife Method	65
Regression Equations	66
Hunter Analysis	71
The Weighted Mean Effect Size From Each Study ..	71
Regression Equations	73
Hedges Analysis	74
The Weighted Mean Effect Size From Each Study ..	74
Regression Equations	77
Comparison of Results Using Different Methods of Analysis	78
Confidence Intervals	78
Regression Equations	79
Analysis of Subsets	80
Differences Due to Method of Analysis	88
V CONCLUSIONS AND DISCUSSION	95
Questions Answered in This Dissertation	95
Discussion	100

	<u>Page</u>
APPENDICES	
A STUDIES USED IN THE META-ANALYSIS	104
B REJECTED STUDIES	117
REFERENCES	122
BIOGRAPHICAL SKETCH	137

LIST OF TABLES

	<u>Page</u>
3-1 Teacher Behaviors Defining the Dimensions of Teacher Clarity	34
3-2 Coding the Dimensions of Teacher Clarity	42
4-1 Characteristics of Studies Producing Low Outliers .	56
4-2 Characteristics of Studies Producing High Outliers	58
4-3 Characteristics of Reduced Data Set	59
4-4 Correlations in Data Set	62
4-5 Tukey's Jackknife Method for Determining the Confidence Interval of the Effect Sizes	67
4-6 Confidence Intervals of All Effect Sizes Using Different Methods	78
4-7 Regression Equation Results Using Different Methods	79
4-8 Confidence Intervals of Subsets of Effect Sizes ...	81
4-9 Differences in the Last Digit in the Results in Table 4-8 From Average Corrected Confidence Intervals Using the Four Methods of Analysis	89
4-10 Summary of Differences From Mean Corrected Confidence Intervals Using the Four Methods of Analysis	94
A-1 Teaching Behaviors and Assumed Dimensions of Teacher Clarity	104
A-2 Characteristics and Results of Studies	113
B-1 Rejected Studies and Reason for Rejection	117

LIST OF FIGURES

	Page
1-1 Factors Affecting the Rate of Transmission of Correct Information	8
1-2 Some Factors Affecting the Rate of Achievement Gain of the Students	9
4-1 Frequency Distribution of the Effect Sizes	55

Abstract of Dissertation Presented to the Graduate School of
the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

THE CORRELATION BETWEEN
TEACHER CLARITY OF COMMUNICATION
AND STUDENT ACHIEVEMENT GAIN:
A META-ANALYSIS

By

FRANK FENDICK

August 1990

Chairman: Dr. James Algina
Major Department: Foundations of Education

The problem was to determine the correlation between teacher clarity and the mean class student learning (achievement gain) in normal public-education classes in English-speaking, industrialized countries. The grade range was assumed to be from Grade 1 through undergraduate study. A normal class was defined as one in which the students are not special in any way and the class runs for a minimum of 6 weeks. Class achievement gain was defined as the mean posttest score minus the mean pretest score on a valid (relevant), reliable test of the subject matter that was taught on the course. Teacher clarity was defined as clarity of (a) organization, (b) explanation, (c) examples

and guided practice, and (d) assessment of student learning. Clarity of speech was regarded as a prerequisite of teacher clarity. Student achievement gain was defined as the class posttest score minus the pretest score or its equivalent. Different methods of meta-analysis were used in order to determine whether they resulted in significantly different results.

It is of practical and theoretical importance to know the relationship between class learning and teacher clarity. It is also important to know how the measured relationship varies with the context of the learning and with the method of analysis.

The correlation between teacher clarity and student achievement gain (effect size) was found to be $.35 \pm .05$. The method of meta-analysis used made no difference. The different dimensions of teacher clarity did not produce significantly different effect sizes. A factor score combining at least two dimensions of teacher clarity had a significantly higher effect size: $.60 \pm .13$.

Larger effect sizes were obtained with (a) student raters rather than observers, (b) experienced rather than inexperienced teachers, and (c) college rather than elementary school teachers. Class size and subject taught made no difference to the effect size.

CHAPTER I
INTRODUCTION

A tale should be judicious, clear, succinct;
The language plain, the incidents well link'd;
Tell not as new what ev'ry body knows;
And, new or old, still hasten to a close.

William Cowper, 1731-1800

The topic of this dissertation is the correlation between teacher clarity of communication and the achievement gain of the students. This was estimated by a meta-analysis of all available studies that could be located.

Teacher Clarity of Communication

Clarity of Speech

Communication between teacher and student cannot occur if the student cannot hear or understand what the teacher is saying. Thus teacher ability to speak loudly enough and in a manner such that the students can comprehend the teacher's speech is a necessary, but not sufficient, condition for communication to occur. Clarity of speech (SP) is regarded as a prerequisite of teacher clarity of communication.

Clarity of Organization

It is assumed in this dissertation that the teacher's task is to assist as many as possible of her or his students to pass an examination at the conclusion of the course (with as high a score as possible). The teacher's first task is to determine the end points of the course and the ground

that must be covered: The teacher must determine where the students are at the beginning and exactly what they must be able to do at the end. She or he must then plan to cover the necessary work to be accomplished in the time available. That is, the teacher must give a pretest (oral or written) that reviews the prerequisites for the course and any topics on the course that the students might already know. The teacher must study past forms of the examination paper in order to determine what the students must do in the posttest, and must then schedule the work to be covered, allowing time for review and test practice. This organization of teaching time can be summed up as (a) determining and stating the objectives of the course, (b) covering the topics that are required by the posttest, and (c) reviewing what has been covered. This organization of teaching time has to take place at the level of the individual topic or lesson as well as for the course as a whole. This is clarity of organization (ORG).

Clarity of Explanation

The teacher must explain the subject matter of the lesson in such a way that it is easy for the students to understand. In order to do this the teacher must (a) explain things simply and make them interesting (otherwise the students will not listen), (b) repeat and stress directions and difficult points, (c) introduce new content in small steps and relate it to content that has been already mastered by the students, and (d) teach at a pace

appropriate to the topic and to the students. This is clarity of explanation (EXP).

Clarity of Examples and Guided Practice

The students will not be able to efficiently answer questions of the type that are on the posttest without practice in doing so. The teacher must (a) demonstrate examples of answering posttest-type questions, (b) answer any questions that the students might have, (c) give the students enough time to practice (in class, for homework, and on practice tests), (d) explain points that have not been answered well and provide standards and rules for satisfactory performance, and (e) provide the students with knowledge on how well they are progressing toward scoring well on the posttest. This is clarity of examples and guided practice (EGP).

Clarity of Assessment of Student Learning

The teacher cannot communicate well without receiving feedback from the students. The teacher checks whether the students are understanding by (a) asking questions during the presentation, (b) encouraging relevant discussion, and (c) checking the students' classwork, homework, and tests. This is clarity of assessment of student learning (clarity of feedback from student to teacher) (ASL).

Summary of the Dimensions of Teacher Clarity

The dimensions of teacher clarity are assumed to be clarity of (a) organization, (b) explanation, (c) examples and guided practice, and (d) assessment of student learning

(feedback from student to teacher). Clarity of speech is assumed to be a prerequisite of teacher clarity.

The Problem Studied in This Dissertation

What is the confidence interval of the correlation between teacher clarity of communication and the achievement gain of the students in the population covered by this study? The confidence interval of the correlation is the range of values that is estimated to have a 95% chance of including the true value.

The population. The population of students and teachers assumed to be covered by this study is all classes in public institutions (Grade 1 through undergraduate) where the education is of the American (European) type and the students or teachers are not selected as being in anyway exceptional.

Extension of the problem. Does the confidence interval change with such factors as (a) the methods of meta-analysis used, (b) the dimension of teacher clarity, (c) grade level, (d) subject taught, (e) any other properties of the situation of teaching and learning, or (f) the analysis and reporting of the correlation?

The Rationale for the Study

The rationale for determining the confidence interval of the correlation between teacher clarity of communication and student achievement gain (and how it varies with various factors) rests on the practical use of such information and on the contribution that it can make to the theory of

teaching. The rationale for determining whether the confidence interval varies with the details of the type of meta-analysis used is to determine the simplest method that can produce a valid confidence interval.

Practical Significance and Objectives of the Study

If we know the correlation (r) between teacher clarity and student achievement gain (student learning), we know the proportion of the variance in achievement gain that is accounted for by variance in teacher clarity (given by $r^2 \times 100\%$). In order to determine the correlation between teacher clarity and student achievement gain, it is necessary to conduct a meta-analysis of the relevant studies. In a meta-analysis the researcher quantifies the results from a number of studies in the form of an average correlation coefficient so that the overall magnitude of the average result can be readily grasped (Gage, 1979).

In this study I set out to answer the following questions:

1. What is the strength of the relationship between teacher clarity and student learning?
2. Do clarity of (a) organization of the lesson (and course), (b) explanation (and speech), (c) examples and guided practice, and (d) assessment of student learning have different relationships to student learning?
3. Do student ratings of teacher clarity have a higher correlation with student learning than observer ratings?

4. Is teacher clarity more predictive of student learning in subjects based on student verbal ability or in those based on numerical ability?

5. Is teacher clarity more predictive at college, at secondary school, or at elementary school? Does the accuracy of prediction vary with grade?

6. Is teacher clarity more predictive in large classes?

7. Does teacher clarity have a stronger relationship with student learning when the teacher is experienced than when she or he is inexperienced?

8. Which factors present in the investigation of relationships between teacher clarity and student learning are likely to result in an inaccurate estimation of the correlation?

9. Do the confidence intervals around the mean correlations obtained in these various circumstances vary significantly with the methods of analysis used? If they do, which method is likely to produce the most valid interval? If they do not, which is the easiest method?

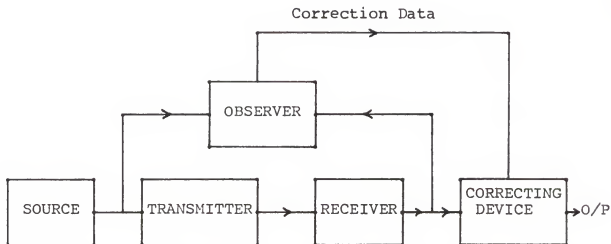
The answers to these questions are clearly important for both the theory and practice of education. From a practical point of view, determining the correlations between teacher clarity and student learning in different circumstances can influence the amount of effort teachers are prepared to exert in order to achieve clarity. It should also influence the emphasis put on this topic by teacher educators.

From the research point of view, a finding that, for example, student rating of teacher clarity is more accurate than observer rating might encourage small groups of teachers to do their own studies of the relationship between teacher clarity (and similar variables) and student achievement in their own situations (grade level, subject area, and type of student). Gage and Needels (1989) stated "In just two of these settings--grade level and subject matter--the need for further process-product research is glaring" (p. 289). Future meta-analyses of hundreds of these small studies would greatly increase our knowledge of important relationships in classroom teaching.

Theoretical Significance of the Study

The relationship between communication variables in the classroom should be subject to an overall theory of communication in any setting. If this study relates the dimensions of teacher clarity to communication theory (also called information theory), and shows that (a) the different dimensions have significantly different correlations with achievement gain and (b) the teacher-behavior variables assumed to define the dimensions produce homogeneous sets of correlations, then the study will have helped to advance our understanding of communication in the classroom.

Communication theory. The basic assumptions of communication theory are illustrated in Figure 1-1.



From "An introduction to information theory: Symbols, signals and noise" by J. R. Pierce, 1980. Copyright 1961 by J. R. Pierce. Adapted by permission.

Figure 1-1. Factors Affecting the Rate of Transmission of Correct Information

The task is to transmit correct information from the source to the output (O/P) of the system at as fast a rate as possible. The rate of transmission depends on (a) the clarity (lack of noise or distortion) of the signal from the transmitter to the receiver, (b) the speed and accuracy with which the observer detects differences between the output of the source and the output of the receiver (detection of errors), and (c) the speed with which the correcting device removes errors.

Communication theory in the classroom. Figure 1-2 shows this model of communication adapted to classroom teaching.

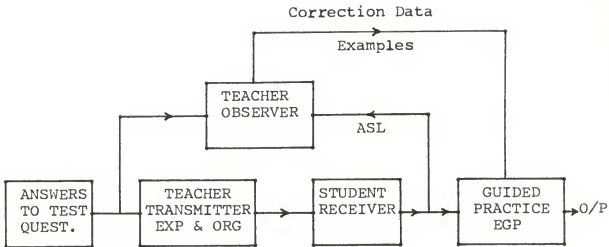


Figure 1-2. Some Factors Affecting the Rate of Achievement Gain of the Students (O/P = output--student performance on posttest).

The rate of transmission of correct information (rate of learning) depends on (a) the clarity of organization and explanation of the signal from the transmitter (teacher) to the receiver (student), (b) the speed and accuracy with which the observer (teacher) detects differences between the output of the source (good answers to test-like questions) and the output of the receiver (detection of errors by the student--assessment of student learning), and (c) the speed with which the correcting device (guided practice) removes errors.

Limitation of This Dissertation

This study can only answer the questions in the preceding section to the extent that the primary research has been done and the results are available in the literature.

Definitions

Definition of Teacher Clarity

Teacher Clarity is assumed to be a measure of the clarity of communication between teacher and students--in both directions. It is assumed to have four dimensions (plus a prerequisite--clarity of speech):

1. Clarity of organization. The teacher must give structure to the lesson (and course). She or he does this by (a) stating objectives and relating them to the course objectives (the topics on the posttest), (b) clearly relating the teaching to the objectives, and (c) reviewing what has been covered in the lesson (and course).
2. Clarity of explanation. The teacher is clear about what he or she explaining and is good at getting the students to understand.
3. Clarity of examples and guided practice--seatwork with the teacher helping as required. The teacher demonstrates on the board examples of the type the students are required to do for seatwork, homework, and tests. The teacher clearly explains as she or he goes through the example what is being done and why. The teacher gradually gets the students to do more of the work themselves until most can make quick and accurate progress without help.
4. Clarity of assessment of student learning--feedback from student to teacher. The teacher cannot hope to achieve clear communication unless she or he studies the students' written, verbal, and nonverbal responses that indicate whether they have understood.

Clarity of speech. In addition to the preceding dimensions, clarity of speech is assumed to be a prerequisite of clarity of explanation. A low score on clarity of speech will necessarily indicate a low score on clarity of explanation; it does not follow, however, that a high score on clarity of speech indicates a high score on clarity of explanation (Cruickshank & Kennedy, 1986). Clarity of explanation is concerned with what the teacher is saying, providing the students can hear and comprehend the language used by the teacher to say it: The teacher speaks loudly enough so that everyone can hear, and her or his accent is not sufficiently different from that of the students to make communication difficult. The teacher speaks with expression and is not monotonous and dull. The teacher's speech is not made difficult to understand by the use of vague terms, mazes (false starts; see, e.g., Hiller, Fisher, & Kaess, 1969), ambiguous pronouns (e.g., the teacher says "he" and the students have no idea to whom the teacher is referring), or continual "uh"s.

Definition of Student Achievement Gain

In an ideal study, the students would be randomly assigned to classes. On the first and last day of the course they would take a relevant, valid, and reliable test of the course objectives. The pretest, a measure of intelligence, present GPA (grade point average), and the reason why the student is taking the course (when optional) would be used to check that random assignment had resulted in there being no significant difference in the students in

each class. If this is found to be true, the class achievement gain is the class mean of the difference between each student's posttest and pretest score (simple gain score). If there is a significant difference in the students, each student's posttest score is predicted from the pretest and precourse measures and the relative influence of the teachers is estimated from the mean for all the students in the class of the difference between a student's actual posttest score and her or his predicted score (residual gain score).

In a real study, random assignment is seldom possible and the pretests and posttests used often do not match the objectives of the course at the appropriate level (Porter, 1985). In this study, I will include all research studies pertaining to the correlation between teacher clarity and class learning unless the posttest used in the study is not a valid measure of the subject matter taught; otherwise (a) the decision to include or not include a study is likely to be subjective, (b) useful information will be discarded, and (c) there will not be a sufficient number of studies to analyze. It will also be necessary to record the measure used to estimate achievement gain in order to test whether this results in a difference in the correlation between teacher clarity and student learning.

Student achievement is assumed to be defined by any achievement measure that (a) is taken by the students in the classes or sections being compared and (b) is given to all classes where the students have not been selected by ability.

(or any variable related to ability) for the different classes unless a pretest is taken so that achievement gain (or its equivalent) can be measured directly.

Outline of the Dissertation

In chapter II, I report and discuss the literature that supports the assumed dimensions of teacher clarity and the low-inference teacher behaviors that are assumed to define those dimensions. I also give (a) an account of some of the problems in conducting meta-analyses, (b) criteria for evaluating them, and (c) some criticisms on how meta-analyses have been used in education. In chapter III, I review the methodology for conducting meta-analyses of correlations according to each of three leading proponents (Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985 and Hedges, 1988; and Hunter, Schmidt, & Jackson, 1982), and detail the procedures that have been used in this analysis. The results are given in chapter IV and the discussion and conclusions comprise chapter V.

CHAPTER II
REVIEW OF THE LITERATURE

Teacher Clarity Literature

The Dimensions of Teacher Clarity

I have assumed that the dimensions of teacher clarity are clarity of (a) organization (ORG), (b) explanation (EXP) with a prerequisite of clarity of speech (SP), (c) examples and guided practice (EGP), and (d) assessment of student learning (ASL). This section presents the evidence for this assumption. The inclusion of ORG, EXP, SP, EGP, or ASL in parentheses indicates information that supports that assumed dimension of teacher clarity.

Solomon, Bezdek, and Rosenberg (1964). These investigators studied 24 teachers of adult evening classes in American government. They observed the teachers, studied audio tapes of lessons, and obtained student evaluations of the lessons. The learning measures were a factual test on the content of the course and a comprehension test not related to the course. There was a negligible correlation between the second test and teacher clarity. A reasonable conclusion is that one should test the course content that one teaches (or teach the course content that is to be tested). Teacher variables were factor analyzed and one of the eight factors was labeled Clarity-Expressiveness vs.

Obscurity-Vagueness. The correlation of this factor with achievement gain in the topics taught was .58. The items that defined the clarity pole of the factor were as follow:

1. Understanding of student statements (ASL);
2. Clear and understandable (EXP);
3. Coherence (EXP); and
4. Well organized (ORG).

Hiller, Fisher, and Kaess (1969). These investigators made frequency counts of 15-minute content-controlled lectures in 12th grade social studies. The five factors used were as follow:

1. Verbal fluency (SP);
2. Optimal information amount (EXP);
3. Knowledge structure cues (ORG);
4. Interest (EXP); and
5. Vagueness (SP).

The correlations between vagueness and student learning were $r(32) = -.59$ and $r(23) = -.48$.

Cruickshank, Myers, and Moenjak (1975) (cited in Cruickshank & Kennedy, 1986). These researchers set out to determine the specific instructional behaviors that students use to discriminate between clear and unclear teachers. They had 1,009 students in grades 6-9 recall their most clear teacher and list the five things that the teacher did. This resulted in the following 12 categories:

1. Providing students with feedback or knowledge of how well they are doing (EGP);

2. Teaching things in a related step-by-step manner (EXP);
3. Orienting and preparing students for what follows (ORG);
4. Providing standards and rules for satisfactory performance (EGP);
5. Using a variety of teaching methods (EXP);
6. Repeating and stressing directions and difficult points (EXP);
7. Demonstrating (EGP);
8. Providing practice (EGP);
9. Adjusting teaching to the learner and the topic (EXP);
10. Providing illustrations and examples (EGP);
11. Communicating so that students understand (EXP);
and
12. Causing students to organize materials in a meaningful way (ORG).

Bush, Kennedy, and Cruickshank (1977). The 110 low-inference behaviors that were detected by Cruickshank et al. (1975) were used to get 1,549 ninth-grade students to rate their most clear and unclear teachers. The 10 behaviors that discriminated best between these two sets of teachers were as follow:

1. Gives the student individual help (EGP);
2. Gives explanations that students understand (EXP);
3. Teaches at a pace appropriate to the topic and the

students (EXP);

4. Takes time when explaining (EXP);
5. Answers student questions (EGP);
6. Stresses difficult points (EXP);
7. Shows students examples of how to do classwork or homework (EGP);
8. Reviews work with students in preparation for a test (ORG);
9. Gives the students enough time to practice (EGP);
and
10. Supports the lesson with specific details (EXP).

Kennedy, Cruickshank, Bush, and Myers (1978). The Bush et al. (1977) items were used with junior high students in Ohio, Tennessee, and Australia. A factor analysis of the results produced the following:

1. Assesses student learning (ASL);
2. Provides student opportunity to practice (EGP);
3. Uses examples (EGP); and
4. Reviews and organizes (ORG).

The 10 most discriminating behavioral statements were as follow:

1. Explains things simply (EXP);
2. Gives explanations the students understand (EXP);
3. Teaches at a pace appropriate to the topic and students (EXP);
4. Stays with the topic until the students understand (EXP);

5. Tries to find out if the students do not understand and repeats things (ASL);
 6. Teaches step-by-step (EXP);
 7. Describes the work to be done and how to do it (EGP);
 8. Asks if the students know what to do and how to do it (EGP);
 9. Repeats things when the students do not understand (EXP); and
 10. Explains something and then works an example (EGP).
- S. Smith (1978) (cited in Cruickshank & Kennedy, 1986).

Observers rated videotapes of 99 community college instructors. Factor analysis resulted in the following factors:

1. Organization (ORG);
2. Makes organization of presentation explicit to students (ORG); and
3. Uses questioning skills, examples (ASL & EGP).

Hines (1981) (cited in Cruickshank & Kennedy, 1986).

The methods used in the preceding studies were duplicated with 573 undergraduate students. The factors produced were as follow:

1. Provides for student understanding and assimilation of instructional content (EXP);
2. Explains/demonstrates how to do the work by the use of examples (EGP); and

3. Structures instruction and instructional content

/presents content in a logical sequence (ORG & EXP).

Cooper and Foy (1967). These researchers analyzed the responses of university students in England and found that the teacher behaviors that the students considered most important were as follow:

1. Presents his material clearly and logically (EXP);
2. Enables the student to understand the basic principles of the subject (EXP);
3. Can be heard clearly (SP);
4. Makes his material intelligibly meaningful (EXP);
5. Adequately covers the ground in the lecture course (ORG);
6. Maintains continuity in the course (ORG);
7. Is constuctive and helpful in his criticism (EGP);
and
8. Shows an expert knowledge of his subject (EXP).

McCaleb and Rosenthal (1983). These researchers factor analyzed both student ratings and observer ratings of the instruction of nine college teaching assistants. The three main factors produced could be called clarity of assessment of student learning (ASL), clarity of examples and guided practice (EGP), and clarity of organization (ORG).

The preceding literature review provides evidence for the assumed teacher clarity dimensions: organization, explanation, examples and guided practice, and assessment of student learning. Clarity of speech is assumed to be a

prerequisite of clarity of explanation rather than a separate dimension. This will now be reviewed.

Clarity of Speech

It is necessary for the teacher to speak loudly enough so that the students can hear. This is not often a problem but, when it is, it is serious. Another problem can be the teacher's accent. This is particularly the case in the teaching of science and math at universities where many of the professors and graduate teaching assistants are foreign. This problem often resolves itself in about two weeks, which seems to be the time it takes for students to get used to an accent. Teacher behaviors that loaded at .35 or above on "verbal fluency" (Bush, Kennedy, & Cruickshank, 1977) were as follow:

1. Speaks grammatically;
2. Speaks with expression; and
3. Speaks so that all the students can hear.

I have not come across any other literature on these two problems except for an article on making effective academic presentations (Renfrow & Impara, 1989) which stated that "verbal behaviors include concerns about pace, pause, pitch, vocal variety, and clarity. Make sure the audience can hear and understand your words" (p. 21).

Hiller, Fisher, and Kaess (1969) defined vagueness as "a psychological construct which refers to the state of mind of the performer who does not sufficiently command the facts or the understanding required for maximally effective

communication" (p. 670). Kounin (1970) found that discontinuities, where the teacher interjects irrelevant content or relevant content at inappropriate times, resulted in loss in lesson momentum.

A typical experimental investigation into the effect of clarity of speech (called teacher clarity by the investigators) is that of Land and L. Smith (1979). They used a "2 (teacher vagueness versus no teacher vagueness) x 2 (teacher mazes versus no teacher mazes) x 2 (additional unexplained content versus no additional unexplained content) experimental design" (p. 55). The subjects (N = 160) were education and psychology undergraduates. They viewed 20-minute videotaped lessons and then completed a 17-item criterion-referenced test. The investigators reported that the results indicated a significant relationship with achievement for vagueness and mazes but not for the inclusion of extra content.

Hiller et al. (1969, quoted in L. Smith & Land, 1981) reported the following to be indicators of vagueness:

Ambiguous designation: Conditions, other, somehow, somewhere, someplace, thing.

Approximation: About, almost approximately, fairly, just about, kind of, most, mostly, much, nearly, pretty (much), somewhat, sort of.

"Bluffing" and recovery: Actually, and so forth, and so on, anyway, as anyone can see, as you know, basically, clearly, frankly, in a nutshell, in essence, in fact, in other words, obviously, of course, so to speak, to make a long story short, to tell the truth, you know, you see.

Error admission: Excuse me, I'm sorry, I guess, I'm not sure.

Indeterminate quantification: A bunch, a couple, a few, a little, a lot, several, some, various.

Multiplicity: Aspect(s), kind(s) of, sort(s) of, type(s) of.

Negated intensifiers: Not all, not many, not very.

Possibility: Chances are, could be, may, maybe, might, perhaps, possibly, seem(s).

Probability: Frequently, generally, in general, normally, often, ordinarily, probably, sometimes, usually. (p. 37)

Chilcoat (1987) included in the vagueness measure the use of (a) pronouns when it is not clear to the students to whom or what the teacher is referring and (b) "I could tell you, but; of course I could; and so on" as part of bluffing. The other indicator of lack of verbal fluency--mazes--was defined by L. Smith (1977) as false starts or halts in speech, redundantly spoken words, and tangles of words. Examples are "will enab . . . will get," and "uh."

Clarity of speech is assumed to comprise (a) speaking in good English loudly enough so that all the students can hear, (b) using few vague terms, and (c) having few false starts or halts in speech.

Clarity of Organization

Brophy (1987) stated that "information presentations are often poorly organized, without advance organizers or other appropriate structure at the beginning, underscoring the main ideas in the middle, or review and summary at the

end" (p. 20). The main components of clarity of organization are (a) providing explanation of objectives at the beginning of the course and lesson, (b) teaching the topics that are covered on the posttest (or specified as the objectives), and (c) summarizing and reviewing at the end of the course or lesson or at the beginning of the following lesson (e.g., Good & Grouws, 1979; Good, Grouws, & Ebmeier, 1983).

Clarity of Explanation

Teacher behaviors that loaded at .35 or above on "explaining" factors produced when ninth graders rated both clear and unclear teachers (Bush, Kennedy, & Cruickshank, 1977) and were not the same behavior expressed in different words (e.g., "Works examples and explains them," and "Explains and then works an example") were as follow:

- Gives explanations that the student understands*
- Teaches at a pace appropriate to the topic and the students*
- Stresses difficult points*
- Uses common words
- Explains new words
- Writes important things on the board
- Answers student questions*
- Teaches one thing at a time
- Repeats enough but not too much
- Reviews work with students in preparation for a test*
- Supports the lesson with specific details*
- Explains by telling a story
- Has students make outlines
- Tells humorous stories when explaining
- Shows movies and explains them afterwards. (pp. 55-56)

Those marked * were reported by Cruickshank and Kennedy (1986) as being prime discriminators between clear and unclear teachers.

The main components of clarity of explanation were assumed to be (a) explaining things simply and interestingly, (b) stressing difficult points, (c) using small steps, and (d) teaching at an appropriate pace.

Examples and Guided Practice

Teacher behaviors that loaded at .35 or above on the Explaining by Examples factors produced when ninth graders rated both clear and unclear teachers (Bush et al., 1977) were as follow:

- Prepares students for what they will be doing next
- Shows students how to do things
- Gives examples and explains them
- Uses common examples
- Shows students examples of how to do classwork or homework*
- Reads the directions with the students
- Asks students before they start to work if they know what to do and how to do it
- Gives the student enough time to practice*
- Gives the students individual help*
- Explains the answers to questions
- Shows the student where he/she is wrong
- Works difficult homework problems selected by students on the board
- Explains in detail what will be on a test
- Takes time to answer students' questions before a test. (pp. 55-56)

Those marked * were reported by Cruickshank and Kennedy (1986) as being prime discriminators between clear and unclear teachers.

The main components of clarity of examples and guided practice are (a) showing students examples of how to do classwork or homework, (b) answering student questions, (c) giving individual help, (d) giving the students enough time to practice, and (e) providing the students with feedback on how well they are doing.

Assessment of Student Learning

Solomon, Bezdek, and Rosenberg (1964) reported that the item with the highest loading on the factor "Obscurity, Vagueness versus Clarity, Expressiveness" was "the teacher's proficiency at receiving the communications of the students" (p. 29). Assessment of student learning is concerned with all the ways in which the teacher learns how well the students are receiving the message the teacher is transmitting; for example, (a) asking questions during the presentation, (b) encouraging relevant discussion, and (c) checking students' work and tests.

With the aid of this literature review, the four assumed dimensions of teacher clarity (plus the prerequisite: clarity of speech) have now been defined in terms of observable low-inference teacher behaviors.

Analysis Literature

General Problems of Meta-analysis

Orwin and Cordray (1984) stated that "meta-analysis is still a relatively new enterprise, and as such it warrants further exploration before conventions regarding proper conduct are adopted" (p. 72). Some of the problems that have been discussed in the literature are as follow:

Apples and oranges. "Criticisms of meta-analysis have primarily revolved around the the issue of 'combining apples and oranges.' That is, combining the results of different studies runs the risk of producing an amalgam that makes no conceptual sense" (Slavin, 1984, p. 9). The reply of the

meta-analysts (e.g., Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Hunter, Schmidt, & Jackson, 1982) is that if sets of studies are different in an important way, then the sets will produce significantly different effect sizes. If interactions between study properties are important, they will be detected by regression equations.

Public availability. An important characteristic of the scientific method is public availability of the data and of the research process. Bullock and Svyantek (1985) suggested that (a) the list of studies used in the analysis, (b) the coding rules used to convert effect-size characteristics into measured variables, and (c) possibly copies of the analyses performed should be publicly available.

Publication bias. Studies published in journals and books tend to be biased toward positive findings (M. Smith, 1980). One must search for other studies in ERIC (Education Resources Information Center) and similar indexes and in Dissertation Abstracts International (Kraemer & Andrews, 1982).

Selection of studies. Care should be taken that selection of studies is not biased by the reviewer's preferences as to what constitutes good methodology. All studies that meet the criteria for inclusion (which must be given) must be included. Studies which fail to meet the criteria should be cited and the reason for exclusion stated (Kraemer & Andrews, 1982).

Treating effect sizes from the same study as independent. M. Smith and Glass (1977) claimed that treating nonindependent data as if they were independent has no consistent impact on the mean effect size. This claim is supported by the reanalysis performed by Landman and Dawes (1982).

Differential attrition. "The important question is whether there is differential attrition in the . . . groups" (Landman & Dawes, 1984, p. 72). This problem is particularly important in college classes: If a larger number of poor students drop one class than drop another, the difference between class mean achievement gains is not a fair comparison of the mean student learning in the classes. A reasonable solution would be to assign, to those students who drop, an achievement gain two class standard deviations below the class mean.

Coding. Detailed decision rules are required for nominal values, for combinations, and for missing data. If estimates are made from published data, the methods of estimation should be explicit. Percentage agreement between independent coders should be given (Bullock & Svyantek, 1985).

Domain of generalization. It is the meta-analyst's right to define the domain of generalization, but it is important that conclusions be limited to that domain (Bullock & Svyantek, 1985).

Criteria for Evaluating Meta-analyses

Bullock and Svyantek (1985) suggested the following criteria for evaluating meta-analyses:

1. Uses a theoretical model as the basis of the meta-analysis;
2. Identifies precisely the domain to be tested;
3. Includes all publicly available studies in the defined content domain;
4. Avoids selecting studies based on criteria of methodological rigor, age of study, or publication status;
5. Publishes or makes available the final list of studies used;
6. Selects and codes variables on theoretical grounds;
7. Provides detailed documentation of the coding scheme including estimation procedures used for missing data;
8. Uses multiple raters to apply the coding scheme and provides a rigorous assessment of interrater reliability;
9. Reports all variables analyzed;
10. Publishes or makes available the data set used in analysis;
11. Considers alternative explanations for the findings obtained;
12. Limits generalization of results to the domain specified by the research;
13. Reports study characteristics in order to

understand the nature and limits of the domain actually analyzed; and

14. Reports the entire study in sufficient detail to allow for direct replication.

Criticisms of Meta-analyses in Education

Slavin (1984) criticized not the concept of meta-analysis but how it had been used in practice in education. His criticisms included the following:

Carlberg and Kavale (1980). These researchers investigated the achievement and social outcomes of placement of exceptional children in special classes rather than in mainstream classes. In most studies the children who were compared were matched only on IQ. Slavin pointed out that there were probably other reasons (such as behavioral difficulties) for placing one student in the special class rather than in the mainstream so that the comparison would inevitably lead to an apparent advantage for mainstreaming.

Kulik and Kulik (1982). These researchers investigated effects of the tracking (class ability grouping) of students by IQ and prior achievement. Slavin stated that (a) in one study no account was taken of a difference of eight IQ points in the groups of students being compared and (b) that in comparing high achievers in high-track classes with high achievers in low-track classes no account was taken of why these high achievers were on different tracks. The mean effect in studies using random assignment of students was zero; thus Kulik and Kulik's claim for the superiority of

tracking on achievement rests solely on studies that are most likely to suffer from selection bias.

Johnson, Johnson, and Maruyama (1983). This analysis investigated the effects of cooperation on relationships between students. Slavin's main complaint was that in the cooperation groups the students were instructed to work together and in the competition groups the students were to work alone. The conclusions of the analysis were (a) that there was more cooperation in the cooperation groups and (b) that groups of students solved problems quicker than did individual students. These conclusions are trivial.

Glass, Cahen, Smith, and Filby (1982). This study purported to study the effects of class size on achievement. In classes of normal size (20-40) there was practically no effect, but when normal-size classes were compared to tutoring (1-3 students) there was a considerable effect. Glass et al. concluded that class size did have a meaningful effect on achievement when this was patently not true in the normal range of classes.

These early meta-analyses show that it is just as important in meta-analysis as it is in other analyses to be certain that the data are really addressing the problem that one should be investigating.

CHAPTER III
REVIEW OF METHODOLOGY

Objective

The objective of a meta-analysis of correlational studies is to determine the best estimate of the confidence interval surrounding the correlation between one variable and another from all available information in the literature. The confidence interval is the value of $\underline{R} \pm d\underline{R}$, where \underline{R} is the point estimate of the population correlation, $d\underline{R}$ is the uncertainty in that estimate, and \pm means plus or minus.

The result of the meta-analysis might be just one confidence interval that covers all the circumstances covered in the studies or, more often, a number of confidence intervals that are valid in specific circumstances. For example, the result of this study might have been that the correlation between teacher clarity and student achievement gain is $.35 \pm .05$, whatever the definition of teacher clarity and whatever achievement is being measured. More likely, a different confidence interval is obtained when teacher clarity is defined as clarity of explanation rather than clarity of organization, or when achievement is based on student verbal ability rather than on student numerical ability.

The purpose of this chapter is to discuss the methods that have been proposed for determining the best values of R (point estimate of correlation) and dR (the uncertainty) and for determining whether one confidence interval (distribution of R) is significantly different from another.

Uncertainty and Variance

The confidence interval is the range of values that we can be 95% confident includes the true value of the population correlation, R . If the confidence interval does not include .00, the value of R is significant at the .05 level. The uncertainty in a mean or correlation is given by twice the standard deviation in the mean (standard error) or correlation provided the value of N (the number of whatever is the unit of analysis) is at least about 20. The variance is the square of the standard deviation. Thus, discussing the reduction in the variance is the same as discussing reduction of the uncertainty or narrowing of the confidence interval.

Summary of Procedure

The meta-analytic procedure consists of (a) defining the problem; (b) finding the studies; (c) describing, classifying, and coding research studies; and (d) analyzing the data. There is no disagreement on how to carry out the first three of these but much disagreement about the techniques of analysis. I will (a) define the problem and describe the methods used for finding and describing the studies and (b) detail the techniques of analysis

recommended by Glass, McGaw, and Smith (1981); Hunter, Schmidt, and Jackson (1982); Hedges and Olkin (1985); and Hedges (1988). The teacher-clarity data have been analyzed by the different techniques in order to determine whether (in this case) they lead to substantively different conclusions.

Defining the Problem

In this dissertation, the problem was to determine the correlation between teacher clarity and the mean class student learning (achievement gain) in normal public-education classes in English-speaking, industrialized countries (e.g., USA, UK, and Australia). The grade range was assumed to be from Grade 1 through undergraduate study (called Grade 13). A normal class was defined as one in which the students are not special in any way and the class runs for a minimum of 6 weeks. Class achievement gain was defined as the mean posttest score minus the mean pretest score on a valid (relevant), reliable test of the subject matter that was taught on the course. Teacher clarity was defined in the ways given in Table 3-1.

The criteria for the inclusion of an effect size (correlation) from a study were (a) the unit of analysis was the class rather than the individual student, (b) a common achievement measure (that covered the content taught) was used across all classes, and (c) data were available to calculate the correlation between the rating of the teacher

Table 3-1. Teacher Behaviors Defining the Dimensions of Teacher Clarity

Clarity of Speech (SP)

(Focus is on absence of inhibitors of communication)

1. Speaks so that all the students can hear
2. Speaks good English without a marked accent
3. Uses few vague terms
4. Speaks with few mazes (false starts or halts in speech, e.g., "uh")

Clarity of Organization (ORG)

(Focus is on objectives, content coverage, and review)

1. States objectives of the course and lesson
2. Covers all the topics on the posttest: Teaches the topics that are specified as the objectives of the course
3. Reviews the lesson at the end of the lesson or at the beginning of the following lesson.
4. Reviews work with students in preparation for a test

Clarity of Explanation (EXP)

(Focus is on the teacher's presentation)

1. Explains things simply and makes them interesting
2. Repeats and stresses directions and difficult points
3. New content is introduced in small steps and is related through ideas held in common by contiguous discourse units (kinetic structure)
4. Teaches at a pace appropriate to the topic and the students

Examples and Guided Practice (EGP)

(Focus is on the teacher's efforts to help the student)

1. Shows students examples of how to do classwork or homework
2. Answers student questions
3. Gives the students individual help
4. Gives the students enough time to practice
5. Explains points that have not been answered well on classwork, homework, and tests
6. Provides standards and rules for satisfactory performance
7. Provides students with feedback or knowledge of how well they are doing

Assessment of Student Learning (ASL)

(Focus is on communication from student to teacher)

1. Asks the students questions during the presentation
 2. Encourages relevant discussion
 3. Checks students' classwork, homework, and tests
-

on at least one dimension of teacher clarity and the class mean achievement gain. Effect sizes based on posttest only were included, but it was recorded whether or not there was evidence of the random assignment of students to classes.

The problem was extended to determine whether the correlation between teacher clarity and class learning depends on such variables as (a) the definition of teacher clarity, (b) the student ability (verbal or quantitative) on which achievement in the subject depends, (c) teacher experience (teacher/professor versus student teacher/teaching assistant), (d) the grade level, (e) normal or experimental class, and/or (f) the characteristics (e.g., quality and year) of the study.

Finding the Studies

The methods used to find the studies for this analysis were (a) tracing back from the references in the studies already located, especially review studies; (b) conducting computer searches of indexes such as ERIC (Education Resources Information Center), Dissertation Abstracts, Psychological Abstracts, and NTIS (National Technical Information Service); (c) supplying a bibliography to researchers in the field and asking them to let me know of any studies I had missed; and (d) manually searching recent editions of likely journals that have not yet been added to the indexes.

Describing, Classifying, and Coding Research Studies

The Dimensions of Teacher Clarity

The concept of teacher clarity of communication in this dissertation is based on communication (information) theory (see chapter I). The basic idea is that the teacher makes a clear, well-organized presentation of the material. The teacher then gets the students to practice answering questions of the type that are on the posttest and provides guidance to the students while they are practicing. Any efficient system must have a feedback mechanism. In this communication system the feedback during the presentation is obtained by the teacher asking the students questions. During practice sessions feedback is obtained by the teacher encouraging student discussion and by the teacher checking the students' classwork, homework, and tests. The teacher then reviews with the students topics and methods that require extra work.

Communication between teacher and student cannot be clear if the student cannot hear or understand the teacher. Thus, lack of clarity of speech is regarded as an inhibitor of teacher clarity. If clarity of speech is a problem, other dimensions of teacher clarity do not get the chance to operate efficiently.

On the basis of this theory and on the basis of the literature reviewed in chapter II, teacher clarity is assumed to have the dimensions given in Table 3-1. Clarity of speech has already been dealt with. Clarity of

organization focuses on statement of objectives, the content coverage (in terms of the topics on the posttest) of the lesson and course, and on review. Clarity of explanation focuses on the teacher's presentation (keep it simple, stress difficult points, use small steps connected to each other, and go at the correct pace). Clarity of examples and guided practice focuses on the teacher's efforts to help the students answer test-like questions by giving them opportunities to practice, showing examples on the board, and by assisting individuals or groups. Assessment of student learning is the feedback from student to teacher and comprises asking the students questions, encouraging relevant discussion, and checking classwork, homework, and tests.

Some teacher-behavior factors that were used in the studies are a mixture of the above dimensions. In this case one must decide whether the factor is predominantly one of the above dimensions or whether it is such a mixture that it is better to categorize it as such and call it teacher skill (SKI).

The following decision-making procedure was found useful in categorizing the teacher behaviors reported in the studies:

Was the teacher behavior concerned with whether

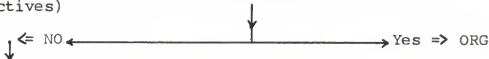
1. the students hear and understand the teacher's speech? (loud enough, accent not too foreign, few vague terms, few false starts, few "uh"s)



2. the teacher was asking a question, listening to answers or discussion, or checking student work?



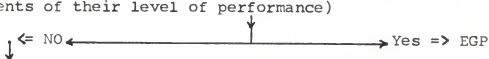
3. the teacher was stating objectives or conducting review? (including relevance of objectives to posttest and relevance of teaching to objectives)



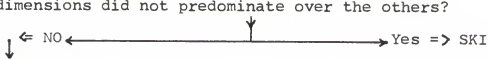
4. the teacher's presentation of the topic was simple, interesting, used small steps that were related to each other, was conducted at the appropriate pace, and emphasized the important points? (the teacher has mastered both the topic and teaching the topic)



5. the teacher was helping the student to perform at the level required by the posttest? (showing by example, answering questions, providing practice time, setting standards, informing students of their level of performance)



6. the teacher was acting in ways covered by more than one of the above dimensions and one of the dimensions did not predominate over the others?



Try again.

The teacher behaviors in the studies were assigned dimensions independently by four coders. The reliability of classification was found to be .76.

Description and Classification

The results of describing and classifying studies are presented in Appendix A. In Table A-1, the information given is an identification number (ID), the name(s) of the author(s) and the year of publication, the teacher-clarity behaviors given in the study, and the assumed dimensions of the behaviors. In Table A-2, after the ID number, the information given is as follows:

1. VER refers to whether the subject taught is classified as achievement being mainly dependent on the student's verbal ability (1) or numerical ability (-1).
2. PUB refers to whether the study was published in a journal or book (1), or is a dissertation or has not been published (-1) (e.g., ERIC microfiche of an address at an annual meeting).
3. STU refers to whether the rating of teacher clarity was made by students (1), by observers (-1), or by both students and observers (0).
4. ACH is an estimate of the validity of the comparison of the achievement gain of one class with another. In order to validly compare the achievement gain of one class with that of another, one should ideally have the same students in each class or at least have random assignment of students to classes. This is often not possible, so an effort is made to compensate for initial differences in students by measuring the difference between the actual achievement of the students and the achievement

that could be predicted from their initial characteristics, that is, a residual gain score. If no account is taken of initial differences, a simple gain score (difference between posttest and pretest) is used. If only a posttest is used (and there is no evidence of the random assignment of students), the validity of comparison of the achievement gain is likely to be low. This is especially so if the achievement measure is of the essay type rather than of the objective type. The worst situation is when the achievement measure is of the essay type and is graded by the class teacher. With this in mind, the following codes are used: Essay test rated by the class teacher (0), posttest only with no random assignment of students (1), simple gain score (2), residual gain score (3), random assignment of students or the same students rate all the teachers (4).

5. REL is the reliability of the teacher-clarity measure.

6. TEX refers to whether the teachers were experienced (1), were student teachers or teaching assistants (-1), or were both (0).

7. WKS is an estimate of the number of weeks between the start of the course and the posttest.

8. GRA is the grade (college = 13, 8.5 = grades 8 and 9).

9. NS is the average number of students in a class.

10. NC is the number of classes in the study.

11. DIM is the assumed dimension of teacher clarity: ORG indicates clarity of organization, EXP indicates clarity of explanation, EGP indicates examples and guided practice, ASL indicates assessment of student learning, SP indicates clarity of speech, and SKI indicates a factor score comprising more than one dimension and no one dimension is dominating the factor.

12. \underline{R} is the correlation between the dimension of teacher clarity and the achievement gain of the class (the average value of all the \underline{r} s reported in the study for the dimension, the grade, and the subject area).

13. \underline{SMR} is the study mean value of \underline{R} obtained by averaging over all the \underline{R} s for the study.

14. \underline{SMZ} is the Fisher \underline{z} equivalent of \underline{SMR} .

Coding Nominal Values

Nominal values have to be coded to use them in regression equations. Is it better to compare the correlation (effect size-- \underline{r}) in one category with the correlation in another (dummy coding) or with the mean correlation over all the studies (effect coding)? I decided to use effect coding. For example, in coding the dimensions of teacher clarity, the coding shown in Table 3-3 was used. If all the variables are coded -1, the dimension is SKI. If X is coded 1 and all the other variables are coded 0, the dimension is EXP, and so on.

Table 3-2. Coding the Dimensions of Teacher Clarity

Dimension	Code				
	X	O	G	A	S
SKI	-1	-1	-1	-1	-1
EXP	1	0	0	0	0
ORG	0	1	0	0	0
EGP	0	0	1	0	0
ASL	0	0	0	1	0
SP	0	0	0	0	1

Techniques of Analysis

Looking at the Data

An indispensable approach to understanding one's results is to study plots of the data and of residuals (Pedhazur, 1982). In this case the frequency distribution of the effect sizes (correlations) was plotted and the effect sizes were also plotted as a function of each of the study variables. The correlations between the variables were also obtained. The plots and correlations were studied to see what relationships were suggested.

The effect sizes were also regressed on all of the study variables. The residuals (difference between predicted and actual value in standard deviation units) were plotted as a function of the predicted value. In approximately normal data, the residuals are randomly scattered about the zero line, and about 95% of them are less than two standard deviations from the line.

The outliers (those data more than two standard deviations from the line) were checked to see if any mistakes could be detected in analyzing those studies. No mistakes were detected, so the data were removed from the data set and the properties of the studies were reported that might explain the difference between the effect size(s) produced by a study (or group of studies) and the majority of the studies. The new data set was regressed on the variables as before, and the residuals were studied for curvature or heteroscedasticity to see if any function of any of the independent variables (e.g., $\log X$, $1/X$, or X^2) was suggested as a good predictor of the effect size.

Problems in Accumulating the Effect Sizes from Each Study

For fully independent effect sizes each study would produce just one effect size, and that effect size would refer to just one dimension of teacher clarity, one grade level, one subject area and so on. In practice the average of all the effect sizes in the study is used or each effect size is treated as independent of the study from which it came. The way in which these problems are treated by Glass, McGaw, and Smith (1981); by Hunter, Schmidt, and Jackson (1982); and by Hedges and Olkin (1985) and Hedges (1988) will now be explained.

Glass, McGaw, and Smith (1981)

Glass et al. (1981) stated that it makes no practical difference whether one accumulates r s, r^2 s or z s so the analyst should do whichever she or he prefers. On the

problem of the nonindependence of the effect sizes they stated

"Studies" cannot be considered the unit of data analysis without aggregating findings above the levels at which many interesting relationships can be studied. . . . There is no simple answer to the question of how many independent units of information exist in the larger data set. . . . Two resolutions of the problem can be envisioned: one risky, the other complex.

The simple (but risky) solution is to regard each finding as independent of the others. The assumption is untrue, but practical. (p. 200)

One complex method suggested by Glass was Tukey's jackknife method. In this method (a) the mean correlation is determined using all the studies and multiplied by the number of studies (K), (b) the means with the effect sizes from each study deleted in turn are multiplied by K-1, and (c) the values in (b) are subtracted from the value in (a) to give K pseudo study correlations from which the uncertainty can be determined.

Hunter, Schmidt, and Jackson (1982)

Hunter et al. considered that it is correct to cumulate rs rather than zs. They state that zs give larger weights to large correlations than to small ones resulting in a positive bias of up to .07 in the mean correlation.

Concerning the problem of using individual correlations or the study average correlation, they stated

If a set of indicator variables is statistically as well as psychologically equivalent . . . , then the ideal cumulation within a study is confirmatory factor analysis with communalities. . . . The average correlation is usually noticeably poorer. If the set of indicators deviates considerably from the unifactor model, then the set of individual correlations should be contributed to the larger cumulation. . . . The unifactor hypothesis cannot even be tested for many

studies. For such studies the choice defaults to the use of individual correlations versus the use of the average correlation. (pp. 122-123)

I chose to average correlations within studies, within teacher-clarity dimensions, and within grades and subject areas. For the overall confidence interval these values were averaged to give a study mean which was used in the cumulation procedure.

The Hunter cumulation procedure is

1. The frequency-weighted mean (frequency is number of classes) and variance of the effect sizes are corrected for (a) sampling error, (b) error of measurement, and (c) range variation.

2. If considerable variance remains after the adjustment for statistical artifacts, the correlations are examined for evidence of moderator variables or are analyzed by subsets. Selected properties that vary across studies are coded and correlated with study r_s . One relies upon theoretical, logical, statistical, and psychometric considerations when possible in deciding what study characteristics to code and how to code them.

3. Correlations among characteristics and the regression of r on study characteristics are computed. The resulting beta weights are interpreted as indicating potential causal effects of true study characteristics on true study R_s .

Rationale of the Hunter and Schmidt method. The correlation between teacher clarity and achievement gain

varies with the circumstances, such as numerical or verbal subject, grade level, and teacher experience. Thus, there is a population distribution of correlations with a mean of \underline{R} and a standard deviation of \underline{S} . Field-study attempts to determine this distribution introduce sampling error, error of measurement, and (possibly) range restriction that increase the standard deviation (or variance = standard deviation squared) and (in the case of measurement error and range restriction) shift the value of the mean. This produces a sample distribution with a mean of \underline{r} and a standard deviation of \underline{s} .

The goal is to estimate the variance introduced by sampling error and subtract it from the observed variance. The mean and variance are also adjusted for the effects of measurement error and (when necessary) range restriction. The result is an estimate of the population distribution. This population can then be tested for homogeneity and, if necessary, split into smaller data sets.

Correction for sampling error. The formula for the expected sampling error variance is $(1 - \underline{R}^2)^2 \times K/N$, where \underline{R} is the mean correlation, K is the number of effect sizes that have been entered into the cumulation, and N is the total number of classes.

Correction of the confidence interval for error of measurement. The mean can be corrected for error of measurement by dividing the obtained mean value of \underline{r} by the square roots of the mean reliabilities of the two measures.

In this case I am interested in the correlation between true teacher clarity (see later) and measured achievement gain, that is, the achievement gain as it is usually measured in practice rather than the achievement gain that would be obtained if it were measured without error. I have, therefore, corrected for error of measurement in teacher clarity only. (True teacher clarity is defined as the mean rating the teacher would receive from an infinite number of students being taught under the same circumstances as those actually being taught.)

Correction for restriction in range. When the range of values of the independent variable (teacher clarity) is not the same as the range for which one wishes to estimate a correlation, a correction can be made using a function of the ratio of the standard deviation of the observed values to the standard deviation of the desired range. In this analysis the observed range of teacher clarity is assumed to be the range that occurs in the population of teachers so that there is no need to make this correction.

Test of homogeneity of R_s . The test statistic is N (number of classes) times the observed variance divided by $(1 - \bar{R}^2)^2$. The test statistic is compared to the 5% value of chi squared with $(K - 1)$ degrees of freedom where K is the number of effect sizes. If the test statistic is smaller than the value of chi squared, the effect sizes are not diverse. Hunter et al. stated that this is strong evidence that there is no true variation across studies

(Hedges and Olkin, 1985 take a different view--see later).

Analysis by subsets. If the effect sizes are found to be diverse, it is necessary to split the data set into sets divided by such variables as (a) verbal or numerical subject, (b) dimension of teacher clarity, or (c) teacher experience, and to repeat the above procedure on these reduced data sets.

Correlations among characteristics and the regression of the values of r on study characteristics. The correlations among characteristics and the regression of \underline{r} on the characteristics tell us more about the dependence of \underline{r} on these variables.

Hedges and Olkin (1985)

Hedges and Olkin showed that the Pearson \underline{r} is a biased estimate of the true correlation with the bias estimated as $-\underline{r}(1 - \underline{r}^2)/2N$. The magnitude of this bias is less than .01 when the value of \underline{r} is about .3 and the value of N is greater than 18. I, therefore, do not need to be concerned with this bias in this analysis (average number of classes per study is 41).

The sampling variance of \underline{r} is estimated by $(1 - \underline{r}^2)^2/N$. In order to make the variance independent of the value of \underline{r} , Hedges and Olkin stated that the \underline{r} s should be converted to Fisher \underline{z} s before cumulation where

$$\underline{z} = .5\log((1 + \underline{r})/(1 - \underline{r}))$$

and the number of degrees of freedom is $(N - 3)$. The lower value of the 95% confidence interval of \underline{Z} (the mean of \underline{z}) is

given by $\underline{Z} - 2/(N - 3)^{1/2}$ and the upper value by $\underline{Z} + 2/(N - 3)^{1/2}$. This confidence interval is then converted to a confidence interval in \underline{R} (mean of \underline{r}) using a table of conversions.

Cumulation procedure. The \underline{r} s are converted to \underline{z} s and the weighted sample mean \underline{Z} found using weights $(NC - 3)$ where NC is the number of classes in the study. This sample mean is used as the estimate of the population mean. The population variance in this mean is given by $1/(N - 3K)$ where N is the total number of classes and K is the number of effect sizes used in the cumulation. Note that this method ignores the observed variance in the \underline{z} s in favor of a formula. This is not very good science: If theory (formula) provides one value and experiment (observation) another, the experimental value is valid until it can be shown that there is something wrong with the experiment.

Measurement error. The value of the new mean is found by dividing by the square root of the reliability of the measurement of teacher clarity.

Test of homogeneity of \underline{z} s. The test statistic is the sum of $(NC - 3)d^2$ where d is $(\underline{Z} - \underline{z})$. The test statistic is compared to the 5% value of chi squared with $(K - 1)$ degrees of freedom where K is the number of effect sizes. Hedges et al. warned the reader not to take this test too seriously: If the number of effect sizes is large, even small variation can produce a significant value of chi-squared. (Hunter takes a different view--see earlier.)

Fitting general linear models to the z s. A weighted least squares procedure is used with the weight specified as $(NC - 3)$. The number of predictors (p) must be less than the number of effect sizes used (K). The chi-squared statistic for testing the model specification is the "error sum of squares." If this value is less than the 5% value of chi squared with $(K - p)$ degrees of freedom, the model is a good fit.

Hedges (1988)

The population variance in z is given by the observed variance minus the expected variance. The expected variance is given by $K/(N - 3K)$ where N is the total number of classes and K is the number of studies. To correct for the unreliability in the measurement of teacher clarity, the variance is divided by the reliability of teacher clarity, or the point estimates of the limits of the 95% confidence interval are divided by the square root of the reliability.

Summary of Analyses Used in This Study

Data inspection. The frequency distribution of the effect sizes and the frequencies of categories were determined. Effect sizes (r s) were plotted and regressed on study characteristics. Trends and residuals were inspected, and outliers were removed from the data set.

Glass et al. (1981). The methods are as follow:

1. Accumulate r s.
2. Find unweighted mean and variance treating each effect size as independent. Assume that this mean and

variance are estimates of the population mean and variance. (Correct for measurement error in teacher clarity.)

3. Find unweighted mean and variance using Tukey's jackknife method. (Correct for measurement error in teacher clarity.)

4. Determine variation of confidence interval with effect-size characteristics by analyzing in subsets and using unweighted regression equations.

Hunter et al. (1982). The methods are as follow:

1. Accumulate rs.

2. Find weighted mean (R) and variance using the average value from each study and weighting by the number of classes--NC.

3. Estimate the population variance by subtracting the estimated sampling error variance-- $(1 - \frac{R^2}{K})^2 \times K/N$, where K is the number of effect-sizes (studies in this case) and N the total number of classes.

4. Correct the mean and population standard deviation for measurement error in teacher clarity by dividing by the square root of the mean reliability reported in the studies.

5. Conduct a test of homogeneity: The test statistic is N (number of classes) times the observed variance divided by $(1 - \frac{R^2}{K})^2$. The test statistic is compared to the 5% value of chi squared with (K - 1) degrees of freedom where K is the number of effect sizes.

6. Determine variation of confidence interval with effect-size characteristics by analyzing in subsets and using weighted--NC regression equations.

Hedges et al. (1985). The methods are as follow:

1. Accumulate \underline{z} s [$\underline{z} = .5\log((1 + \underline{r})/(1 - \underline{r}))$].

2. Find weighted $(NC - 3)$ mean using the average \underline{z} from each study. Ignore the observed variance and assume the variance in the mean is given by $1/(N - 3K)$, where N is the total number of classes and K is the number of studies. Assume that these are the population values of mean and variance.

3. Correct the mean and standard deviation for measurement error in teacher clarity.

4. Conduct a test of homogeneity: The test statistic is the sum of $(NC - 3)d^2$ where d is $(\underline{Z} - \underline{z})$ and \underline{Z} is the mean uncorrected for measuring error. The test statistic is compared to the 5% value of chi-squared with $(K - 1)$ degrees of freedom where K is the number of effect sizes.

5. Determine variation of confidence interval with effect-size characteristics by analyzing in subsets and using weighted $(NC - 3)$ regression equations. The number of predictors (p) must be less than the number of effect sizes used (K).

Hedges (1988). Method: Same as in Hedges et al. (1985) except that $K/(N - 3K)$ is used as the estimated sampling variance and is subtracted from the observed variance to estimate the population variance.

CHAPTER IV RESULTS AND ANALYSES

Results

The results for the studies that met the criteria for inclusion in the meta-analysis (see chapter 3) are shown in Table A-1 of Appendix A. There are 47 studies, of which 8 report only a reliability of the measure of teacher clarity. There are 39 studies reporting 110 values of the mean correlation between one dimension teacher clarity and class achievement gain. Where a study reported the results separately for verbal subjects (VER) and numerical subjects (NUM), the effect sizes for each are given separately. This is also the case when a study reported separate results for different grades. Within each of these categories the mean of all the correlations in a particular teacher-clarity dimension is the effect size for that dimension. The study mean over all the effect sizes is also reported for each study.

Studies that were judged to have failed to meet at least one of the criteria for inclusion are given in Table B-1 of Appendix B. This table gives the reason for rejecting the study.

Frequency Distribution

The frequency distribution of 109 effect sizes (one value of $-.73$ not included) is shown in Figure 4-1. The fact that the distribution is bell-shaped (rather than looking as though the left side has been cut off) indicates that it is not likely that the publicly available studies tend to be only those with significant positive correlations.

Removal of Outliers

The values of r were regressed on all the study characteristics using the SAS (Statistical Analysis System) GLM (General Linear Models) program (see later) and the residuals plotted against the ID number of the study. A residual is the difference between the actual value of r and its value predicted by the regression model. Ten values of r were found to have residuals greater than two standard deviations of the residuals and were removed from the data set.

Low outliers. The characteristics of the studies that produced low outliers are given in Table 4-1. There does not seem to be anything in common with these seven studies:

In four the subject is numerical, two verbal, and one both.

Three different dimensions of teacher clarity--EXP, ASL, and EGP--are represented.

Two were published in a journal or book, four were in ERIC, and one was a dissertation.

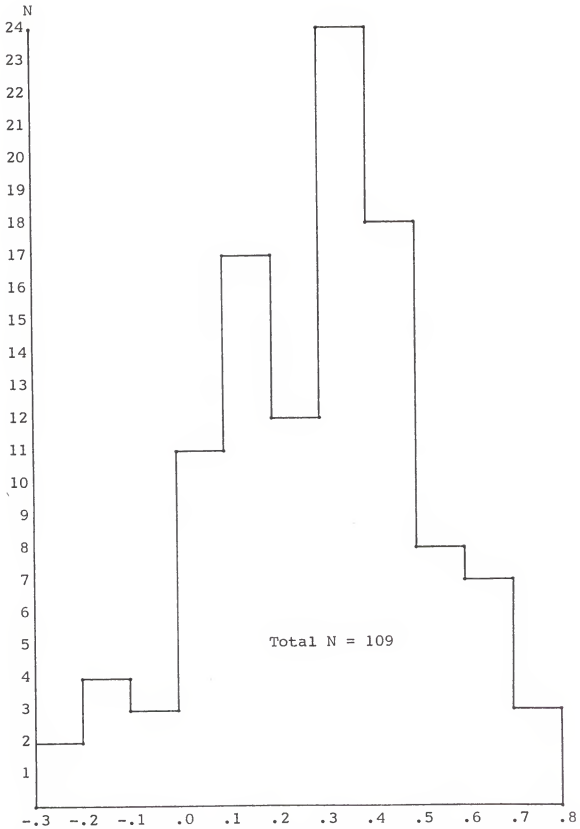


Figure 4-1. Frequency Distribution of the Effect Sizes

Table 4-1. Characteristics of Studies Producing Low Outliers

ID	VER	PUB	STU	ACH	REL	TEX	WKS	GRA	NS	NC	DIM	<u>R</u>	<u>SMR</u>	<u>SMZ</u>
8	-1	1	-1	2	.	1	30	5	30	36	EXP	-.18	.12	.12
22	1	1	-1	3	.	1	2	4	25	16	ASL	-.19	.10	.10
23	1	-1	1	0	.	-1	10	13	17	31	EXP	-.19	-.04	-.04
30	-1	-1	-1	3	.75	1	30	2.5	25	75	ASL	-.14	.12	.12
34	0	-1	-1	3	.45	1	20	5.5	30	26	ASL	-.73	-.19	-.19
35	-1	-1	-1	2.5	.45	1	30	5	20	41	EGP	-.2	.18	.18
42	-1	-1	1	3	.89	-1	10	13	30	36	EXP	-.25	-.25	.26

Note. ID: identification number of study; VER: 1 = subject based on students' verbal ability, -1 = subject based on students' numerical ability, 0 = both subject areas; PUB: 1 = study published in a journal or book, -1 = dissertation or ERIC; STU: 1 = student rating of teacher clarity, -1 = observer rating of teacher clarity, 0 = rating by both students and observers; ACH: 0 = essay test rated by the class teacher, 1 = posttest only with no random assignment of students, 2 = simple gain score, 3 = residual gain score, 4 = random assignment of students or the same students rate all the teachers; REL = reliability of the teacher-clarity measure; TEX: 1 = experienced teachers, -1 = student teachers or teaching assistants; WKS = weeks between the start of the course and the posttest; GRA = grade (college = 13, 8.5 = grades 8 and 9); NS = average number of students in a class; NC = number of classes in the study; DIM: ORG = clarity of organization, EXP = clarity of explanation, EGP = examples and guided practice, ASL = assessment of student learning, SP = clarity of speech, SKI = a factor score comprising more than one dimension and no one dimension is dominating the factor; R = the correlation between the dimension of teacher clarity and the achievement gain of the class (the average value of all the rs reported in the study for the dimension, the grade, and the subject area); SMR = the study mean value of R averaged over all the Rs for the study; SMZ = the Fisher z equivalent of SMR.

Four used residual gain measures of achievement, one simple gain, one both residual and simple gain, and one posttest only.

In five the teachers were experienced and in the other two were learners.

Five of the courses were normal (6 weeks or more) and one experimental.

Grade level varied from 4 through 13 (college).

Both the number of students (NS) and the number of classes (NC) were quite high (greater than 16) in all the studies.

The nearest suggestion to something in common is that teacher clarity was rated by observers rather than by students in five of the studies, and there is something not quite satisfactory in the two studies that were rated by students:

In Benton (1976) the achievement measure was an essay test that was graded by the class teacher. This achievement score is likely to be less valid than most of the estimates of class achievement gain because of the lower reliability of essays compared to objective measures of achievement.

In Hazelton (1980) 19 factors were produced from the answers to 108 questions by 1,102 students in 36 classes. Nineteen is a suspiciously high number of factors, and 108 questions is a high number of questions to expect students to answer conscientiously.

High outliers. The characteristics of the studies producing high outliers are given in Table 4-2. The points in common with the studies is that they were published in a journal and teacher clarity was rated by college students.

Table 4-2. Characteristics of Studies Producing High Outliers

ID	VER	PUB	STU	ACH	REL	TEX	WKS	GRA	NS	NC	DIM	<u>R</u>	<u>SMR</u>	<u>SMZ</u>
11	1	1	1	2	.	-1	15	13	26	17	EXP	.81	.46	.50
43*-1	1	1	4	.	.	1	15	13	78	20	ORG	.77	.73	.93
44 -1	1	1	3	.	.	1	10	13	35	13	EXP	.79	.73	.93

Note. To interpret the heading abbreviations see the note below Table 4-1.

* The same students rated all 10 teachers in 20 subject areas.

Reliability of Dimensions of Teacher Clarity

Four teachers completed the task of classifying teacher behaviors into the dimensions of teacher clarity. The percentage of agreement with my classifications varied from 64% to 93% with a mean of 76%. Many measures only achieve a reliability of about .8, so I judged this value to be high enough to consider the classification as reasonably reliable. Most disagreement occurred in classifying factors (comprising a number of different teacher behaviors) as either SKI (more than one dimension) or as being dominated by a particular dimension.

Characteristics of the Reduced Data Set

The characteristics of the data set are shown in Table 4-3. The prototypical study (a) was conducted in the

Table 4-3. Characteristics of Reduced Data Set

Characteristic	(Total number of $r = 100$)	N
<u>Educational setting</u>		
Elementary school (Grades 1 - 6)	48
Secondary school (Grades 7 - 12)	15
College (Grade 13)	37
<u>Decade published</u>		
60s	5
70s	78
80s	17
<u>Number of students in class--NS (Mean = 27)</u>		
Large NS (30 and above)	42
Small NS (less than 30)	58
<u>Number of classes in study--NC (Mean = 41. Total = 98--2 values not known)</u>		
Large NC (40 and above)	35
Small NC (less than 40)	63
<u>Number of Normal classes (course lasting at least 6 weeks with the regular teacher)</u>		
Normal	80
Experimental (sometimes just one lesson)	20
<u>Verbal or numerical ability</u>		
Verbal	48
Numerical	47
Both	5
<u>Studies Published or not (not = ERIC or dissertation)</u>		
Published	81
Not published	19
<u>Teacher-clarity raters</u>		
Students	40
Observers	60

continued

Table 4-3 continued

Characteristic	(Total number of $r = 100$)	N
<u>Validity of comparison of class achievement gain</u>		
Posttest only with essay test graded by class teacher (coded 0)		1
Posttest only (no random entry of students to class) (coded 1)		9
Simple gain (posttest - pretest) (coded 2)		12
Correlations reported for both simple gain and residual gain (coded 2.5)		3
Residual gain (difference between actual and expected gain) (coded 3)		45
Evidence of random entry of students to classes or same students rating different teachers (coded 4)		8
Validity not known		22
<u>Number of studies reporting reliability of teacher clarity</u> (Total number = 24. Mean reliability = .78)		
Reliability approximately .5		2
Reliability approximately .8		13
Reliability approximately .9		9
<u>Experienced teachers or learners</u> (learners = teaching assistants or student teachers)		
Experienced teachers		65
Learners		28
Both		7
<u>Number of weeks between start of teaching and the posttest</u> (Mean = 17 weeks)		
Less than 4 weeks		20
4 - 11 weeks		6
12 - 15 weeks		34
16 - 27 weeks		3
28 - 30 weeks		37
<u>Dimensions of teacher clarity</u>		
Assessment of student learning--ASL		21
Examples and guided practice--EGP		17
Clarity of explanation--EXP		31
Clarity of organization--ORG		21
Clarity of speech--SP		4
Overall rating of teacher clarity--SKI		6

elementary school or college rather than secondary school, (b) was published in the 1970s, (c) had 27 students per class and 41 classes per study, (d) was of a normal class (lasting a semester or a year) with the regular (experienced) teacher, anequally likely to depend on numerical ability as on verbal ability, (e) was more likely to use observers rather than students as raters of teacher clarity and the reliability of the rating was about .8, (f) was most likely to use a measure of achievement gain (simple or residual) rather than to use posttest only and/or random assignment of students to classes, and (g) investigated any of the four dimensions of teacher clarity (ASL, EGP, EXP, ORG) rather than the prerequisite of teacher clarity (clarity of speech--SP) or the overall rating of teacher clarity (SKI).

Relationships Between Characteristics

Correlations between the characteristics are shown in Table 4-4. The correlations between teacher clarity and student achievement gain (effect size-- r) are shown to increase with grade (higher in college than in school), when the studies are published in journals or books (rather than ERIC or dissertations), and when students do the rating rather than observers. Among those (24) effect sizes from studies that report the reliability of the teacher-clarity measure, the effect size increases as the reliability of the teacher-clarity measure increases.

Table 4-4. Correlations in Data Set

**YR	VER	PUB	STU	ACH N=79	REL N=24	TEX	WKS	GRA	NS N=98	NC	
VER	-28*										
PUB	26*	-15									
STU	-09	00	25*								
ACH	-26*	-05	-14	-08							
REL	36	04	70*	31	-40*						
TEX	05	-17	-18	-53*	33*	-48*					
WKS	15	-13	-21*	-22*	06	-53*	41*				
GRA	03	05	31*	76*	-14	43*	-61*	-58*			
NS	00	-09	09	04	26*	-49*	33*	26*	-09		
NC	-20*	22*	-14	17	12	-02	-05	00	22*	-06	
<u>r</u>	11	-07	35*	27*	-19	66*	-08	-11	31*	13	-14

Note. Decimal point omitted, N = 100 when not given.

**YR = year study reported, VER = subject based on verbal ability, PUB = published in journal or book, STU = teacher clarity rating by students, ACH = validity of achievement gain, REL = reliability of teacher clarity measure, TEX = experienced teacher, WKS = weeks of course, GRA = grade, NS = number of students in class, NC = number of classes in study, r = correlation between teacher clarity and student achievement gain.

*Significant ($p = .05$ or less).

Other significant correlations between variables show that

1. More recent studies have (a) tended to study subjects based on the students' numerical ability rather than their verbal ability, (b) tended to be published in journals or books, (c) used measures of achievement gain that are assumed to be of lower validity, and (d) used fewer classes in the study.

2. Studies of verbal subjects tend to use a higher number of classes in the study.

3. Published studies tend to (a) use students as raters, (b) have a higher reliability of measurement of teacher clarity, and (c) be at the college level. (At the college level the courses are about 15 weeks. At school the courses are about 30 weeks. Thus, if the majority of published studies are at the college level (Grade 13) there will be a negative correlation between PUB and WKS. There is also a positive correlation between PUB and GRA. These two correlations indicate that the study was at the college level.)

4. Students are more likely to do the teacher-clarity rating (a) when the teachers are learners and (b) at the college level.

5. The assumed validity of the measure of achievement gain tends to be (a) negatively related to the reliability of the measure of teacher clarity, (b) be higher when the teachers are experienced than when they are learners, and

(c) be higher in classes with a large number of students.

6. The reported reliability (only 24 studies) is higher (a) when the teachers are learners, (b) at the college level, and (c) when the number of students in the class is low.

7. Experienced teachers tend to (a) be studied in school rather than college and (b) have a larger number of students than learner teachers.

8. The length of a course is (a) longer at school than at college and (b) longer when the number of students in the class is higher (normal classes tend to have a higher number of students than experimental classes that sometimes consist of only one lecture).

9. At higher grade levels the number of classes used in the study tends to be larger.

Glass Analysis

The data used in the analyses are shown in Table A-2 in Appendix A (R in the table is referred to as r in the analyses as R is used as the mean). In the following analysis the methods of Glass et al. (1981) were used.

Treating Each Effect Size as Independent

The unweighted mean for the 100 values of r in the reduced data set (i.e., after the removal of the outliers) was .30 with a standard deviation of .19. The standard error (standard deviation of the mean--standard deviation divided by the square root of the number of values) was

therefore .02 and the uncertainty in the mean is twice this value, .04. The 95% confidence interval of the population value of the mean correlation between all dimensions of teacher clarity and the achievement gain of the students is therefore between .26 and .34.

Glass does not recommend correcting for the unreliability of the teacher clarity measure, but in order to make a comparison with the results by Hunter's method and Hedges's method the correction was made. The mean reliability of the teacher clarity measure is .78. The square root of this is .88. Dividing the original limits of the confidence interval by this value, the population correlation is estimated to be between .30 and .39.

What was the effect of dropping the outliers? The mean of all 110 effect sizes was .28 with a standard deviation of .25. This gives a standard error of .25 divided by the square root of 118, that is .024, and an uncertainty of .05. The confidence interval of the population mean using the full data set is between .23 and .33. Correcting for unreliability in the teacher clarity measure, it is between .27 and .38. This does not vary appreciably from the confidence interval obtained using the reduced data set.

Using Tukey's Jackknife Method

In this method, pseudo-values of \bar{r} for each study are calculated by (a) obtaining the mean value of \bar{r} from all 38 studies (using all the values of \bar{r}) and multiplying this mean by 38; (b) obtaining 38 means of \bar{r} from 38 studies,

dropping all the effect sizes from one study at a time, and multiplying each mean by 37; and (c) subtracting each result in (b) from the value obtained in (a) in order to obtain 38 pseudo-values that represent the effect of dropping each study from the data set. The population confidence interval is then estimated from the mean and uncertainty in the mean calculated using these pseudo-values.

The value for (a) was found to be $38 \times .29990 = 11.396$ and the values for (b) and (c) were as shown in Table 4.5. This leads to a mean of .31 and an uncertainty of .04 so the population mean is estimated to be between .27 and .35. Correcting for the unreliability in the measure of teacher clarity, this becomes .31 through .40. Thus, using the jackknife method merely raises the confidence interval by .01 above the confidence interval obtained from using the original effect sizes.

Regression Equations

Model. The model used in the regression equation was

$$\underline{r} = \text{YR VER PUB STU ACH TEX WKS GRA NS X G O A S}$$

where \underline{r} = correlation between teacher clarity and
achievement gain

YR = year of report

VER = subject based on verbal ability

PUB = published in book or journal

STU = student rating of teacher clarity

ACH = validity of comparison of achievement gain

TEX = Experienced teacher

Table 4-5. Tukey's Jackknife Method for Determining the Confidence Interval of the Effect Sizes

ID	Values of \bar{r}	(Mean)	(b)*	(c)**
1	.46 .06 .38	(.30)	11.248	.312
2	.40 .45 .22 .32 .49 .53 .27 .50	(.38)	10.915	.645
3	.34 .32 .06 .42 .33 .44 .17 .37	(.32)	11.248	.312
4	.07 .13	(.11)	11.386	.174
5	.37 .39 .34	(.39)	11.174	.386
6	.20	(.20)	11.285	.275
7	.31	(.31)	11.248	.312
8	.36 .15	(.26)	11.359	.201
9	.08 .30 .10	(.16)	11.433	.127
10	.30 .48	(.39)	11.211	.349
11	.42 .21 .40	(.34)	11.211	.349
12	.34 .18	(.27)	11.285	.275
13	.59 .29 .36	(.45)	11.036	.524
14	.22 .40 .36	(.33)	11.248	.312
15	.14	(.14)	11.322	.238
16	.47 .51	(.49)	11.137	.423
17	.10 .61 .21	(.31)	11.248	.312
18	.58	(.58)	11.174	.386
19	.41 .21 .53 .01	(.29)	11.285	.275
20	.07	(.07)	11.359	.201
21	.26 .32	(.29)	11.285	.275
22	.07 .11 -.06 .44	(.14)	11.507	.053
23	-.19 .11	(-.04)	11.322	.238
27	.30 .24 .00	(.18)	11.396	.164
29	.49	(.49)	11.174	.386
30	.20 .02 .27 .18 .17 -.14 .04 .19	(.12)	11.692	.132
31	.54	(.54)	11.174	.386
32	.68 .61	(.65)	10.989	.571
33	.11	(.11)	11.322	.238
34	-.03 -.03	(-.03)	11.507	.053
35	.30 .33	(.32)	11.248	.312
37	.07 .13	(.11)	11.211	.349
39	.47	(.47)	11.211	.349
40	.14 .01 .14	(.10)	11.470	.090
41	.43	(.43)	11.211	.349
43	.69	(.69)	11.100	.460
44	.67	(.67)	11.137	.423
45	.68 .32 .67	(.56)	10.952	.608

Mean .31
SD .14
SE .02

Note. (a) = 38 x mean of all 38 values = 38 x .30421 = 11.560
*(b) = 37 x mean of all values except that study
**(c) = (a) - (b)

WKS = Length of course

GRA = Grade (college = 13)

NS = Number of students in class

X = 1 when dimension of teacher clarity = EXP

G = 1 when dimension of teacher clarity = EGP

O = 1 when dimension of teacher clarity = ORG

A = 1 when dimension of teacher clarity = ASL

S = 1 when dimension of teacher clarity = SP

and X, G, O, A, S = - 1 when dimension = SKI.

Only 76 values of \underline{r} were used because of missing values. The mean of these 76 values was .29 with an uncertainty of .04. Thus the mean of the 76 values is only .01 less than that obtained earlier with 100 values. The model accounted for 46% of the variance in \underline{r} .

Significant variables. The variables with significant (.05) Type I sums of squares (assumes variable is entered in the order given in the equation) were PUB, STU, and GRA. These variables were also significantly correlated with the effect size (Table 4-4). The variables with significant Type III sums of squares (assumes variable is entered in last in the equation--which is the same as testing the significance of the regression coefficient) were ACH (negative), TEX, and GRA.

The fact that PUB and STU were no longer significant when entered last indicates that the difference in effect size reported in published studies (or those with student raters) compared to those reported in unpublished studies

(or with observer raters) was due to the correlations between PUB (STU) and other variables. For example, Table 4-4 shows that published studies correlate with grade .31 (STU with GRA .75) and that grade correlates with effect size .31. Thus, the higher effect size in published studies is due in part to the fact that higher effect sizes are obtained in a college setting (see Table 4-8) and that a higher percentage of published studies are at the college level rather than the elementary school level.

Varying the model. The model was reduced by removing the dimensions of teacher clarity. This reduced the variance in r accounted for by 9% to 37%. This difference is not significant. The only significant regression coefficient was GRA. ACH ($p = .08$) and TEX ($p = .057$) were nearly significant. PUB ($p = .89$) was far from significant.

The number of classes (NC) used in obtaining the value of r was added to the previous model. This increased the variance accounted for by less than 1%, and the regression coefficient was not significant. Thus, the number of classes in the study was not a meaningful source of variance. GRA was still the only significant regression coefficient.

Only 76 effect sizes were used in the above analyses because of 22 missing values of ACH (16 in Berliner & Tikunoff, 1977, 3 in Bryson, 1974, and 3 in Bourke, 1985) and 2 missing values of NS (Centra, 1977). The missing values of ACH were set at 2.5 (the median value), and the

original model was then run with 98 observations. The mean of the 98 is .30, which is only .01 more than that obtained earlier with 76 values. The model accounted for 36% of the variance in \bar{r} instead of the 46% accounted for with 76 observations. The variables with significant (.05) Type I sums of squares were still PUB, STU, and GRA. The variables with significant Type III sums of squares (and regression coefficients) were ACH (negative), TEX, and PUB. Note that with the addition of the 22 effect sizes PUB becomes significant and GRA is no longer significant.

When NC (number of classes that contributed to the effect size) was added to the 98-observations model, the variance in \bar{r} accounted for increased by 3% to 39%. The significant regression coefficients with this model were ACH, TEX, and PUB as in the model without NC, but also included GRA and NC.

Summary. The significant regression coefficients with the various models and observations were as follows:

1. When 76 observations and the teacher clarity dimensions were used, ACH (the validity of comparison of achievement gain between classes) was negative and significant. TEX (experienced teacher) and GRA (grade) were positive and significant.

2. With 76 observations and no teacher clarity dimensions only GRA was significant and it was positive. There was no significant change when NC (number of classes contributing to the effect size) was added to the model.

3. When 98 observations and the teacher clarity dimensions were used, ACH (negative), TEX, and PUB (published in journal or book) were significant. GRA ($p = .14$) was not significant.

4. When NC was added to the model in 3, NC and GRA ($p = .03$) were added to the significant variables: ACH, TEX, and PUB.

I will report how these results compare with those obtained by the Hunter and Hedges methods before analyzing by subsets.

Hunter Analysis

The methods of Hunter et al. (1982) were used in the analysis that follows.

The Weighted Mean Effect Size from Each Study

Calculations.

$$K = \text{number of studies} = 38$$

$$N = \text{total number of classes} = 1699$$

$$\text{Observed variance} = .0325$$

$$\begin{aligned} \text{Estimated sampling variance} &= (1 - \underline{R}^2)^2 \times K/N \\ &= (1 - .30^2)^2 \times 38/1699 \\ &= .0185 \end{aligned}$$

$$\begin{aligned} \text{Estimated population variance} &= .0325 - .0185 \\ &= .0140 \end{aligned}$$

$$\text{Estimated population SD} = .12$$

$$\text{Estimated population SE} = .12/38^{1/2} = .02$$

$$\text{Uncertainty} = .04$$

$$\begin{aligned} & \text{95\% confidence interval of population mean} \\ & = .30 \text{ plus or minus } .04 \\ & = .26 \Rightarrow .34 \\ & \text{Corrected mean} = .30 / .78^{1/2} = .34 \end{aligned}$$

$$\begin{aligned} & \text{Corrected confidence interval} \\ & = .30 \Rightarrow .39 \end{aligned}$$

Test of homogeneity:

$$\begin{aligned} \text{Observed variance} \times N &= .0325 \times 1699 \\ &= 55.22 \end{aligned}$$

$$\begin{aligned} \text{Expected variance} &= (1 - \underline{R})^2 \\ &= (1 - .30)^2 = .8281 \end{aligned}$$

$$\text{Test statistic} = 55.22 / .8281 = 66.7$$

$$\text{Chi-square with 37 degrees of freedom} = 52$$

Therefore the effect sizes are not homogeneous.

Commentary. The values of \underline{r} are weighted by the number of classes which contribute to the correlation on the assumption that the larger the number of classes the more valid the value. The weighted mean for the 38 mean values of \underline{r} in each study is .30 with a standard deviation of .18 as shown in the calculations. The observed variance (SD squared) is .0325. The estimated variance due to sampling error is .0185, leaving an estimated population variance of .0140. This gives a population standard deviation of .12, a standard error of .02, and an uncertainty in the mean of .04. The 95% confidence interval of the population value of the mean correlation between all dimensions of teacher clarity and the achievement gain of the students is

therefore between .26 and .34. This is exactly the same interval that was obtained using the Glass method. Correcting for the unreliability of teacher clarity results in an interval of .30 through .39.

To test the uncorrected result for homogeneity, the test statistic (ratio of the observed variance to the expected variance if all sample correlations were estimates of a single population correlation) is compared to the .05 value of chi-square using the number of studies minus one (37) as the number of degrees of freedom. The test statistic (66.7) is greater than the chi-square value (52) so the effect sizes are not homogeneous, and it is therefore necessary to use regression equations and analyze by subsets.

Regression Equations

The model used in the regression equation was

$$\underline{r} = \text{YR VER PUB STU ACH TEX WKS GRA NS X G O A S}$$

where the variables are the same as in the Glass method but the effect size (\underline{r}) is weighted by the value of NC (the number of classes contributing to \underline{r}) on the assumption that the larger the number of classes the more valid the value. The significant variables were determined by using generalized least squares implemented on the WEIGHT = NC option of SAS GLM.

The model accounted for 54% of the variance in \underline{r} compared to 46% using the Glass unweighted method. Setting the missing values of ACH to 2.5 and using 98 observations

reduced the variance accounted for to 43%. The variables with significant regression coefficients in the model with 76 observations were ACH (negative), TEX, and GRA. With 98 observations, GRA ($p = .42$) was no longer significant; ACH and TEX remained significant. PUB was not significant in either case.

Hedges Analysis

The methods of Hedges and Olkin (1985) and Hedges (1988) were used in the following analysis.

The Weighted Mean Effect Size From Each Study

Hedges and Olkin (1985) calculations.

$$K = \text{number of studies} = 38$$

$$N = \text{total number of classes} = 1699$$

$$\text{Observed variance} = .0421$$

$$\text{Mean } \underline{Z} = .33 \text{ (Mean } \underline{R} = .32)$$

$$\begin{aligned} \text{Estimated population variance in } \underline{Z} &= 1/(N - 3K) = 1/1585 \\ &= 1/(N - 3K) = 1/1585 \\ &= .0066 \end{aligned}$$

$$\text{Estimated population } \underline{SE} = .0066^{1/2} = .025$$

$$\text{Uncertainty} = .05$$

$$\begin{aligned} \text{95\% confidence interval of population mean } \underline{Z} \\ &= .33 \text{ plus or minus } .05 \\ &= .28 \Rightarrow .38 \\ &(\underline{R} = .27 \Rightarrow .36) \end{aligned}$$

$$\text{Corrected mean } \underline{Z} = .33/.88 = .38$$

$$\text{Corrected mean } \underline{R} = .36$$

$$\begin{aligned} \text{Corrected confidence interval of } \underline{Z} \\ &= .32 \Rightarrow .43 \text{ (} \underline{R} = .29 \Rightarrow .41) \end{aligned}$$

Hedges and Olkin (1985) commentary. The weights used are NZ where $NZ = NC - 3$ as that is the number of degrees of freedom in \underline{z} when the class is the unit of analysis. The weighted mean for the 38 mean values of \underline{z} in each study is .33 ($\underline{r} = .32$). The estimated standard error is .025 giving an uncertainty of .05. The 95% confidence interval for population mean of \underline{z} is from .28 through .38 ($\underline{r} = .27$ through .36). This compares with .26 through .34 obtained by the previous two methods.

Hedges (1988) calculations.

$$K = \text{number of studies} = 38$$

$$N = \text{total number of classes} = 1699$$

$$\text{Observed variance} = .0421$$

$$\text{Estimated sampling variance} = K/(N - 3K)$$

$$= 38/1585$$

$$= .0324$$

$$\text{Estimated population variance in } \underline{z} = .0421 - .0324$$

$$= .0181$$

$$\text{Estimated population } \underline{SD} = .0181^{1/2} = .13$$

$$\text{Estimated population } \underline{SE} = .13/38^{1/2} = .022$$

$$\text{Uncertainty} = .04$$

$$\text{95\% confidence interval of population mean } Z$$

$$= .33 \text{ plus or minus } .04$$

$$= .29 \Rightarrow .37$$

$$(\underline{R} = .28 \Rightarrow .36)$$

$$\text{Corrected mean } \underline{Z} = .38 \text{ (Mean } \underline{R} = .36)$$

$$\text{Corrected confidence interval of } \underline{Z}$$

$$= .33 \Rightarrow .42 \text{ (} \underline{R} = .32 \Rightarrow .40).$$

Hedges (1988) commentary. The mean of \underline{z} is .33 with an observed variance of .0421. The estimated variance due to sampling error is .0324 leaving an estimated population variance of .0181. This gives a population standard deviation of .13, a standard error of .02, and an uncertainty in the mean of .04. The 95% confidence interval of the population value of the mean \underline{z} is therefore between .29 and .37, which corresponds to a confidence interval for the mean \underline{r} of .28 through .36. This is practically the same interval that was obtained using all the other methods. Correcting for the unreliability of teacher clarity results in an interval of .32 through .40.

Test of homogeneity.

$$\begin{aligned}\text{Test statistic} &= \text{observed variance} \times (N - 3K) \\ &= .0421 \times 1585 = 66.7\end{aligned}$$

Chi-square with 37 degrees of freedom = 52

Therefore the effect sizes are not homogeneous.

To test the uncorrected result for homogeneity the test statistic (ratio of the observed variance to the expected variance if all sample correlations were estimates of a single population correlation) is compared to the .05 value of chi squared using the number of studies minus one (37) as the number of degrees of freedom. The test statistic (66.7) is greater than the chi-squared value (52) so the effect sizes are not homogeneous, and it is therefore necessary to use regression equations and analyse by subsets.

Regression Equations

The model used in the regression equation was

$$\underline{z} = \text{YR VER PUB STU ACH TEX WKS GRA NS X G O A S}$$

where \underline{r} = correlation between teacher clarity and
achievement gain

$$\underline{z} = .5 \times \log((1 + \underline{r})/(1 - \underline{r}))$$

YR = year of report

VER = subject based on verbal ability

PUB = published in book or journal

STU = student rating of teacher clarity

ACH = validity of comparison of achievement gain

TEX = Experienced teacher

WKS = Length of course

GRA = Grade (college = 13)

NS = Number of students in class

X = 1 when dimension of teacher clarity = EXP

G = 1 when dimension of teacher clarity = EGP

O = 1 when dimension of teacher clarity = ORG

A = 1 when dimension of teacher clarity = ASL

S = 1 when dimension of teacher clarity = SP

and X, G, O, A, S = - 1 when dimension = SKI.

Each value of \underline{z} was weighted by the value of NZ (the number of classes contributing to \underline{r} minus three, that is, the number of degrees of freedom in \underline{z}). The significant variables were determined by using generalized least squares implemented by the WEIGHT = NZ option of SAS GLM. The number of observations used was 76 out of the 100 due to

missing values of ACH and NS. The model accounted for 57% of the variance in \bar{r} compared to 54% using the Hunter method and 46% using the Glass unweighted method. Setting the missing values of ACH to 2.5 and using 98 observations reduced the variance accounted for was to 46%.

The variables with significant regression coefficients in the model with 76 observations were ACH (negative), TEX, and GRA. With 98 observations, GRA ($p = .34$) was no longer significant; ACH and TEX remained significant. PUB was not significant in either case.

Comparison of Results Using Different Methods of Analysis

Confidence Intervals

Table 4-6. Confidence Intervals of All Effect Sizes Using Different Methods

Method (N = 100, K = 38)**	Uncorrected		Corrected*	
	Mean	Interval	Mean	Interval
<u>Glass et al. (1981)</u>				
Mean of 100 values of \bar{r}				
Unweighted mean	.30	.26--.34	.34	.30--.39
Tukey's Jackknife				
Mean of 38 pseudo study values of \bar{r}	.31	.27--.35	.35	.31--.40
<u>Hunter et al. (1982)</u>				
Mean of 38 study values of \bar{r}				
Variance = observed - formula	.30	.26--.34	.34	.30--.39
<u>Hedges & Olkin (1985)</u>				
Mean of 38 study values of \bar{z}				
Variance by formula	.32	.27--.36	.36	.31--.40
<u>Hedges (1988)</u>				
Mean of 38 study values of \bar{z}				
Variance = observed - formula	.32	.28--.36	.36	.32--.40

Note. *Corrected for unreliability in teacher clarity.
 **N = number of effect sizes, K = number of studies.

Table 4-6 shows that all the different methods essentially give the same confidence interval for the effect size.

Regression Equations

Table 4-7 shows the results of using different models, different numbers of observations, and different weights in the regression analyses.

Table 4-7. Regression Equation Results
Using Different Methods

Model	Weight	N	R ² %	Sig.	Reg. Coeff.
<u>Glass et al. (1981)</u>					
$\bar{r} = YR--S$	1	76	46	ACH(-)	TEX GRA
$\bar{r} = YR--NS$ No TC*	1	76	37		GRA
$\bar{r} = YR--NS+NC$ No TC	1	76	37		GRA
$\bar{r} = YR--S$	1	98	36	ACH(-)	TEX PUB
$\bar{r} = YR--S +NC$	1	98	39	ACH(-)	TEX GRA PUB NC
<u>Hunter et al. (1982)</u>					
$\bar{r} = YR--S$	NC	76	54	ACH(-)	TEX GRA
$\bar{r} = YR--S$	NC	98	43	ACH(-)	TEX
<u>Hedges & Olkin (1985)</u>					
$\bar{z} = YR--S$	NC-3	76	57	ACH(-)	TEX GRA
$\bar{z} = YR--S$	NC-3	98	46	ACH(-)	TEX

Note. *TC = the dimensions of teacher clarity

Significant regression coefficients. ACH (the validity of the comparison of achievement gain) and TEX (experienced teacher) were significant in all cases except when the dimensions of teacher clarity were dropped from the model. Thus we can be fairly certain that (a) the correlation between teacher clarity and achievement gain (i.e., effect size) decreases as the validity of the achievement gain comparison increases, and (b) the effect is greater with

experienced teachers than it is with learners (student teachers and assistant teachers).

GRA (grade) is significant most times: It is likely that teacher clarity is more important at the higher grades than it is in the first years of elementary school. PUB (study published in a journal or book) is significant only in Glass analysis with 98 observations so the large difference in the mean of r for published studies--.38 and the mean for unpublished studies--.21 (see Table 4-8) is largely due to the fairly high correlation between PUB and such variables as GRA and STU (student raters) which are positively correlated with the effect size (see Table 4-4).

Analysis of Subsets

The methods of Glass et al. (1981)--GL, Hunter et al. (1982)--HU, Hedges and Olkin (1985)--HO, and Hedges (1988)--HE were used in the following analyses. Refer to the values corrected for unreliability in the measure of teacher clarity in Table 4-8.

Educational Setting--GRA. The overall mean effect sizes using the means obtained in all analyses (HO and HE count as one result as these analyses obtain the mean by the same method) are (a) elementary school (Grades 1 - 6) .26, (b) secondary school (Grades 7 - 12) .30, and (c) college (Grade 13) .42. All analyses except for HO find the difference between elementary school and college significant (the confidence intervals do not overlap). HE also finds a significant difference between secondary school and college (For HU this difference is nearly significant as the

Table 4-8. Confidence Intervals of subsets of Effect Sizes

Characteristic	N	Anal*	Uncorrected		Corrected**	
			Mean	Interval	Mean	Interval
<u>Educational setting</u>						
Elementary school ...	48	GL	.25	.20--.30	.28	.23--.34
		HU	.22	.19--.25	.25	.22--.28
		HO	.22	.18--.26	.25	.20--.39
		HE	.22	.18--.26	.25	.20--.39
Secondary school	15	GL	.30	.21--.39	.34	.24--.44
		HU	.25	.21--.29	.28	.24--.33
		HO	.25	.21--.29	.28	.24--.33
		HE	.25	.21--.29	.28	.24--.33
College	37	GL	.38	.32--.44	.43	.36--.50
		HU	.35	.29--.41	.40	.33--.47
		HO	.35	.29--.41	.40	.33--.47
		HE	.35	.29--.41	.40	.33--.47
<u>Decade published</u>						
60s	5	GL	.37	.14--.60	.42	.16--.68
		HU	.29	.14--.44	.33	.16--.50
		HO	.29	.14--.44	.33	.16--.50
		HE	.29	.14--.44	.33	.16--.50
70s	78	GL	.30	.26--.34	.34	.30--.39
		HU	.29	.25--.33	.33	.28--.38
		HU	.29	.25--.33	.33	.28--.38
		HU	.29	.25--.33	.33	.28--.38
80s	17	GL	.32	.22--.42	.36	.25--.48
		HU	.28	.24--.32	.32	.27--.36
		HO	.29	.24--.32	.33	.27--.36
		HE	.29	.24--.32	.33	.27--.36
<u>Studies Published or not</u> (not = ERIC or dissertation)						
Published	81	GL	.33	.29--.37	.38	.33--.42
		HU	.32	.28--.36	.36	.32--.41
		HO	.32	.28--.36	.36	.32--.41
		HE	.32	.28--.36	.36	.32--.41
Not published	19	GL	.19	.13--.25	.22	.15--.28
		HU	.18	.14--.22	.20	.16--.25
		HO	.18	.14--.22	.20	.16--.25
		HE	.18	.14--.22	.20	.16--.25

continued

Table 4-8 continued

Characteristic	N	Anal*	Uncorrected		Corrected	
			Mean	Interval	Mean	Interval
<u>Number of students in class--NS</u>						
(Mean = 27)						
Large NS (30 and above) 42		GL	.34	.26--.38	.39	.30--.43
		HU	.39	.35--.43	.44	.34--.49
		HO	.39	.35--.43	.44	.34--.49
		HE	.39	.35--.43	.44	.34--.49
Small NS (less than 30) 58		GL	.28	.23--.33	.32	.26--.38
		HU	.23	.19--.27	.26	.22--.31
		HO	.23	.19--.27	.26	.22--.31
		HE	.23	.19--.27	.26	.22--.31
<u>Number of classes in study--NC</u>						
(Mean = 41)						
Large NC (40 and above) 35		GL	.29	.23--.35	.33	.26--.40
		HU	.27	.22--.32	.31	.25--.36
		HO	.27	.22--.32	.31	.25--.36
		HE	.27	.22--.32	.31	.25--.36
Small NC (less than 40) 63		GL	.31	.26--.36	.35	.30--.41
		HU	.30	.27--.33	.34	.31--.38
		HO	.30	.26--.36	.34	.30--.41
		HE	.30	.26--.36	.34	.30--.41
<u>Normal classes--NOR</u>						
(course lasting at least 6 weeks with the regular teacher)						
Normal	80	GL	.30	.26--.34	.34	.30--.39
		HU	.28	.24--.32	.32	.27--.36
		HO	.30	.26--.34	.34	.30--.39
		HE	.30	.26--.34	.34	.30--.39
Experimental	20	GL	.31	.22--.40	.35	.25--.45
		HU	.30	.25--.35	.34	.28--.40
		HO	.31	.22--.40	.35	.25--.45
		HE	.31	.22--.40	.35	.25--.45
<u>Verbal or numerical ability--VER</u>						
Verbal	48	GL	.34	.29--.39	.39	.33--.44
		HU	.29	.25--.34	.33	.28--.39
		HO	.29	.25--.34	.33	.28--.39
		HE	.29	.25--.34	.33	.28--.39
Numerical	47	GL	.30	.25--.35	.34	.28--.40
		HU	.31	.27--.35	.35	.31--.40
		HO	.31	.27--.35	.35	.31--.40
		HE	.31	.27--.35	.35	.31--.40

continued

Table 4-8 continued

Characteristic	N	Anal*	Uncorrected		Corrected	
			Mean	Interval	Mean	Interval
<u>Validity of comparison of class achievement gain--ACH</u>						
Posttest only (no random entry of students to class) (coded 1)	9	GL	.49	.38--.60	.56	.43--.68
		HU	.54	.51--.57	.61	.58--.65
		HO	.55	.45--.62	.63	.51--.70
		HE	.55	.52--.57	.63	.59--.65
Simple gain (posttest - pretest) (coded 2)	12	GL	.30	.23--.37	.34	.26--.42
		HU	.29	.26--.32	.33	.30--.36
		HO	.29	.16--.42	.33	.18--.48
		HE	.29	.26--.32	.33	.30--.36
Residual gain (difference between actual and expected gain) (coded 3)	45	GL	.25	.19--.31	.28	.22--.35
		HU	.22	.19--.25	.25	.22--.28
		HO	.22	.17--.27	.25	.19--.31
		HE	.22	.19--.25	.25	.22--.28
Evidence of random entry of students to classes or same students rating different teachers (coded 4)	8	GL	.37	.20--.54	.42	.23--.61
		HU	.35	.25--.45	.40	.28--.51
		HO	.36	.16--.62	.41	.18--.70
		HE	.36	.29--.41	.41	.33--.47
Validity not known	22	GL	.33	.27--.39	.38	.31--.44
		HU	.29	.26--.32	.33	.30--.36
		HO	.29	.20--.38	.33	.23--.43
		HE	.29	.26--.32	.33	.30--.36
<u>Experienced teachers learners--TEX</u> (learners = teaching assistants or student teachers)						
Experienced teachers	65	GL	.29	.25--.31	.33	.28--.35
		HU	.27	.23--.31	.31	.26--.35
		HO	.28	.24--.30	.32	.27--.34
		HE	.28	.25--.31	.32	.27--.35
Learners	28	GL	.32	.25--.39	.36	.28--.44
		HU	.25	.21--.29	.28	.24--.33
		HO	.25	.18--.32	.28	.20--.36
		HE	.25	.20--.30	.28	.23--.34
Both	7	GL	.36	.19--.53	.41	.22--.60
		HU	.43	.32--.55	.49	.36--.63
		HO	.45	.36--.54	.51	.41--.61
		HE	.45	.36--.53	.51	.41--.60

continued

Table 4-8 continued

Characteristic	N	Anal*	Uncorrected		Corrected	
			Mean	Interval	Mean	Interval
<u>Teacher-clarity raters--STU</u>						
Students	40	GL	.36	.30--.43	.40	.33--.48
		HU	.34	.29--.39	.39	.33--.44
		HO	.36	.31--.41	.41	.35--.47
		HE	.36	.30--.42	.41	.34--.48
Observers	60	GL	.27	.22--.32	.31	.25--.36
		HU	.24	.20--.28	.27	.23--.32
		HO	.24	.19--.29	.27	.22--.33
		HE	.24	.21--.27	.27	.24--.31
<u>Dimensions of teacher clarity</u>						
ASL	21	GL	.26	.18--.34	.30	.21--.39
		HU	.29	.23--.35	.33	.26--.40
		HO	.30	.22--.38	.34	.25--.43
		HE	.30	.23--.36	.34	.26--.41
EGP	17	GL	.22	.14--.28	.25	.16--.32
		HU	.19	.16--.22	.22	.18--.25
		HO	.20	.13--.27	.23	.15--.31
		HE	.20	.17--.23	.23	.19--.26
EXP	31	GL	.33	.27--.39	.38	.31--.44
		HU	.29	.24--.34	.33	.27--.39
		HO	.30	.21--.39	.34	.24--.44
		HE	.30	.26--.34	.34	.30--.39
ORG	21	GL	.31	.23--.39	.35	.26--.44
		HU	.26	.21--.31	.30	.24--.35
		HO	.26	.18--.34	.30	.20--.39
		HE	.26	.20--.32	.30	.24--.36
SP	4	GL	.36	.04--.68	.41	.05--.77
		HU	.39	.13--.55	.44	.15--.63
		HO	.40	.21--.59	.45	.24--.67
		HE	.40	.23--.46	.45	.26--.52
Overall rating SKI..	6	GL	.54	.40--.68	.61	.45--.77
		HU	.51	.44--.58	.58	.50--.66
		HO	.51	.40--.62	.58	.45--.70
		HE	.51	.45--.57	.58	.51--.65

Note. **Corrected for unreliability in teacher clarity.
 *Method of analysis: GL = Glass, McGaw, & Smith, 1981; HU = Hunter, Schmidt, & Jackson, 1982; HO = Hedges & Olkin, 1985; HE = Hedges, 1988.

confidence intervals just touch at .33). The unweighted mean (GL) is higher than the weighted means (HU, HO, & HE) in all cases.

Decade published--YR. There are no significant differences. The overall mean in all decades is .34. The only result worth noting is that the unweighted mean for the 60s (GL = .42) obtained when there are very few results (5) is a long way from the overall mean. This suggests that using the weighted mean may produce more stable results than using the unweighted mean.

Study published in journal or book--PUB. The overall mean for published studies is .37, and for ERIC documents and dissertations it is .21. This difference is significant by all methods of analysis.

Number of students in class--NS. The overall mean for large classes (30 or more students) is .32, and that for small classes .29. This difference is not significant by any method of analysis as the confidence intervals all have a range of about .12.

Number of classes in the study--NC. The overall mean for large studies (40 or more classes) is .32 and that for small classes .34. This difference is not significant by any method of analysis.

Normal classes--NOR. The overall mean for normal classes is .33 and that for experimental classes .35. This difference is negligible.

Verbal or numerical ability--VER. The overall mean for both verbal and numerical ability is about .35. Note the value of .39 obtained using the unweighted mean (GL) even though there are 48 observations. This value is .06 above those obtained using weighted means, and might be another indication of the instability of the unweighted mean.

Validity of comparison of class achievement gain--ACH. When only a posttest was given and there was no evidence of random assignment of students to classes, the overall mean effect size is .60. When simple gain (a difference score) is used, the mean is .33. When residual gain is used, the mean is .26. When there is evidence of random assignment of students to classes or the same students rate different teachers, the mean is .41. All methods of analysis show that the posttest-only studies result in a significantly higher value than do the gain-score studies. The HU and HE methods give the simple-gain studies a significantly higher mean than that for the residual-gain studies. The GL and HO methods give wide confidence intervals (.11 - .30) for both these sets of studies with the result that the difference in the means is not significant. There are only eight random-entry type effects, so the confidence interval using all methods is very wide. The high mean with posttest-only (coded 1) explains the negative regression coefficient obtained in the regression equations.

Experienced teachers or learners--TEX. The overall mean with experienced teachers is .32 and that with learners

.31. So why was the regression coefficient significant? The explanation lies in the fact that TEX is correlated with GRA (grade) $-.61$, and GRA is correlated with effect size $.31$, which results in a small negative ($-.08$) correlation between TEX and effect size (in line with the slightly smaller effect size for experienced teachers) when unweighted (GL) means are used. Thus, the significant regression coefficient indicates that there is closer relationship between teacher clarity and student achievement gain for experienced teachers than there is for learners after the effect of other variables (such as grade level) have been taken into account.

Teacher-clarity raters--STU. When the rating of teacher clarity was made by students, the mean effect size was $.40$; by observers, it was $.28$. The two means were significantly different for all methods of analysis except for the Glass method. The main reason the regression coefficient for STU was not significant was the $.76$ correlation between STU and GRA (student rating takes place mostly at the college level). Thus, after GRA has been entered into the equation, there is very little independent variance due to STU.

Dimensions of teacher clarity. The overall rating of teacher clarity (SKI, which includes at least two of the dimensions of teacher clarity) produced a significantly higher mean effect size ($.60$) than did the other dimensions. There were only four results for SP (clarity of speech) and

these were spread over a wide positive range so nothing can be said about this dimension--except that the correlation with achievement gain is positive. The HE (Hedges, 1988) method produced a narrower confidence interval than the other methods. Considering the other four dimensions, EGP (examples and guided practice) had the lowest effect size--.23, but this was not significantly different from that for the highest, EXP--.35. The other two dimensions ASL (assessment of student learning) and ORG (clarity of organization) both produced a mean value of .32.

Differences Due to Method of Analysis

Table 4-9 shows the differences between the effect size when a particular method is used and the average value using all methods of analysis. For example the first value, +2, for GL indicates that the mean for GL (.28--see the corrected value of the mean in Table 4-8) is .02 higher than the average value $((.28 + .25 + .25)/3 = .26)$, using the values of the mean obtained from GL, HU, and HO. (In the case of the mean only, HE is not used in the average as the mean using HE is obtained in exactly the same way as the mean using HO.) The next three values for GL show that (a) the low end of the confidence interval is .02 higher than the average value using all methods, (b) the high end of the confidence interval is .01 lower than the average value, and (c) the width of the confidence interval is .03 narrower than the average value.

Table 4-9. Differences in the Last Digit in the Results in Table 4-8 From Average Corrected Confidence Intervals Using the Four Methods of Analysis

Characteristic	N	Anal*	Mean	Interval	Width
<u>Educational setting</u>					
Elementary school ...	48	GL	+2	+2--	-1 -3
		HU	-1	+1--	-7 -8
		HO	-1	-1--	+4 +5
		HE		-1--	+4 +5
Secondary school	15	GL	+4	0--	+8 +8
		HU	-2	0--	-3 -1
		HO	-2	0--	-3 -1
		HE		0--	-3 -1
College	37	GL	+2	+2--	+2 0
		HU	-1	-1--	-1 0
		HO	-1	-1--	-1 0
		HE		-1--	-1 0
<u>Decade published</u>					
60s	5	GL	+6	0--	+13 +7
		HU	-3	0--	-5 -2
		HO	-3	0--	-5 -2
		HE		0--	-5 -2
70s	78	GL	+1	+1--	+1 0
		HU	0	-1--	0 -1
		HU	0	-1--	0 -1
		HU		-1--	0 -1
80s	17	GL	+2	-1--	+9 +10
		HU	-2	+1--	-3 -4
		HO	-1	+1--	-3 -4
		HE		+1--	-3 -4
<u>Studies Published or not</u>					
Published	81	GL	+1	+1--	+1 0
		HU	-1	0--	0 0
		HO	-1	0--	0 0
		HE		0--	0 0
Not published	19	GL	+1	-1--	+2 +3
		HU	-1	0--	-1 -1
		HO	-1	0--	-1 -1
		HE		0--	-1 -1

continued

Table 4-9 continued

Characteristic	N	Anal	Mean	Interval	Width
<u>Number of students in class--NS</u>					
Large NS (30 and above)	42	GL	-3	-3-- -5	-2
		HU	+2	+1-- +1	0
		HO	+2	+1-- +1	0
		HE		+1-- +1	0
Small NS (less than 30)	58	GL	+4	+3-- +5	+2
		HU	-2	-1-- -2	-1
		HO	-2	-1-- -2	-1
		HE		-1-- -2	-1
<u>Number of classes in study--NC</u>					
Large NC (40 and above)	35	GL	+1	+1-- +3	+2
		HU	-1	0-- -1	-1
		HO	-1	0-- -1	-1
		HE		0-- -1	-1
Small NC (less than 40)	63	GL	+1	0-- +1	+1
		HU	-1	+1-- -1	-2
		HO	-1	0-- +1	+1
		HE	-1	0-- +1	+1
<u>Normal classes--NOR</u>					
Normal	80	GL	+1	+1-- +1	0
		HU	-1	-2-- -2	0
		HO	+1	+1-- +1	0
		HE		+1-- +1	0
Experimental	20	GL	0	-1-- +1	+2
		HU	-1	+2-- -4	-6
		HO	0	-1-- +1	+2
		HE		-1-- +1	+2
<u>Verbal or numerical ability--VER</u>					
Verbal	48	GL	+4	+4-- +4	0
		HU	-2	-1-- -1	0
		HO	-2	-1-- -1	0
		HE		-1-- -1	0
Numerical	47	GL	-1	-2-- 0	+2
		HU	0	+1-- 0	-1
		HO	0	+1-- 0	-1
		HE		+1-- 0	-1

continued

Table 4-9 continued

Characteristic	N	Anal	Mean	Interval	Width
<u>Validity of comparison of class achievement gain-ACH</u>					
Posttest only					
(coded 1)	9	GL	-4	-10-- +1	+11
		HU	+1	+5-- -2	-7
		HO	+3	-2-- +3	+5
		HE		+6-- -2	-8
Simple gain					
(coded 2)	12	GL	+1	0-- +1	+1
		HU	0	+4-- -5	-9
		HO	0	-8-- +7	+15
		HE		+4-- -5	-9
Residual gain					
(coded 3)	45	GL	+2	+1-- +4	+3
		HU	-1	+1-- -3	-4
		HO	-1	-2-- 0	-2
		HE		+1-- -3	-4
Evidence of random entry of students to classes					
(coded 4) 8		GL	+1	-3-- +4	+7
		HU	-1	+2-- -6	-8
		HO	0	-8- +13	+21
		HE		+7- -10	-17
Validity not known ..					
	22	GL	+3	+2-- +4	+2
		HU	-2	+1-- -4	-5
		HO	-2	-6-- +3	+9
		HE		+1-- -4	-5
<u>Experienced teachers or learners--TEX</u>					
Experienced teachers					
	65	GL	+1	+1-- 0	-1
		HU	-1	-1-- 0	+1
		HO	0	0-- -1	-1
		HE		0-- 0	0
Learners					
	28	GL	+5	+4-- +5	+1
		HU	-3	0-- -4	-4
		HO	-3	-4-- -1	+3
		HE		-1-- -3	-2
Both					
	7	GL	-6	-13-- -1	-12
		HU	+2	+1-- +1	0
		HO	+4	+6-- 0	-6
		HE		+6-- -1	-7

continued

Table 4-9 continued

Characteristic	N	Anal	Mean	Interval	Width
<u>Teacher-clarity raters--STU</u>					
Students	40	GL	0	-1-- +1	+2
		HU	-1	-1-- -3	-2
		HO	+1	+1-- 0	-1
		HE		0-- +1	+1
Observers	60	GL	+3	+1-- +3	+2
		HU	-1	-1-- -1	0
		HO	-1	-2-- 0	+2
		HE		0-- -2	-2
<u>Dimensions of teacher clarity</u>					
ASL	21	GL	-2	-4-- -2	+2
		HU	+1	+1-- -1	-2
		HO	+2	0-- +2	+2
		HE		+1-- 0	-1
EGP	17	GL	+2	-1-- +3	+4
		HU	0	+1-- -4	-5
		HO	-1	-2-- +2	+4
		HE		+2-- -3	-5
EXP	31	GL	+3	+3-- +3	0
		HU	-2	-1-- -2	-1
		HO	-1	-4-- +3	+7
		HE		+2-- -2	-4
ORG	21	GL	+3	+2-- +3	+1
		HU	-2	0-- -4	-4
		HO	-2	-4-- 0	+4
		HE		0-- -3	-3
SP	4	GL	-2	-13- +12	+25
		HU	+1	-3-- -2	+1
		HO	+2	+6-- +2	-4
		HE		+8- -13	-21
SKI	6	GL	+2	-2-- +7	+9
		HU	-1	+3-- -4	-7
		HO	-1	-2-- 0	+2
		HE		+3-- -5	-8

Note. *Method of analysis: GL = Glass, McGaw, & Smith (1981); HU = Hunter, Schmidt, & Jackson (1982); HO = Hedges & Olkin (1985); HE = Hedges (1988).

In Table 4-10 these values are averaged over the 32 sets of results and the standard deviation is given. Thus for these sets of results the mean effect size using the GL method is on average .012 higher than the average mean effect size obtained using all methods of analysis. The standard deviation in the effect size is .026. This value multiplied by two and reduced to one significant figure is .05. The difference between the mean effect size obtained by the GL method and the mean effect size obtained by averaging the results from all methods therefore varies from about $.01 - .05 = -.04$ through $.01 + .05 = .06$ (the actual variation is from $-.06$ through $.06$).

The differences in the mean due to method have the same magnitude (.01) for all methods. The GL method produces a mean (with these data) about .02 higher than that produced by the other methods. For example, if the GL method produces a mean effect size of .32, then the other methods are likely to produce .30. The variation (due to method) in the value of the mean by the GL method ($\pm .05$) is about twice as much as that produced by the other methods (.03 and .02).

The differences in the width of the confidence interval due to method have almost the same magnitude (.03) for all methods. The GL and HO methods produce a confidence interval about .06 wider than that produced by the HU and HE methods. This is because in the latter methods the estimated sampling variance is subtracted from the observed

variance. The variation (due to method) in the width of the confidence interval (+/- .11 and .09) is greater in the methods accumulating zs (HO and HE) than it is (+/- .06) in the methods accumulating rs (GL and HU).

Table 4-10. Summary of Differences From Mean Corrected Confidence Intervals Using the Four Methods of Analysis

	N	Anal*	Mean	From	To	Width
Mean	32	GL	1.2	-0.8	3.0	2.8
Standard deviation			(2.6)	(1.3)	(2.0)	(2.8)
Mean		HU	-0.5	0.4	-2.3	-2.6
Standard deviation			(1.3)	(1.6)	(2.0)	(2.8)
Mean		HO	-0.5	-1.1	0.8	1.9
Standard deviation			(1.1)	(1.9)	(3.8)	(5.6)
Mean of means		HE	-0.5	1.1	-2.0	-3.1
Standard deviation			(1.1)	(1.1)	(3.0)	(4.6)

Reducing to one significant figure and reintroducing the decimal point						
Mean	32	GL	.01	-.01	.03	.03
2 x Standard deviation			(.05)	(.03)	(.04)	(.06)
Mean		HU	-.01	.00	-.02	-.03
2 x Standard deviation			(.03)	(.03)	(.04)	(.06)
Mean		HO	-.01	-.01	.01	.02
2 x Standard deviation			(.02)	(.04)	(.08)	(.11)
Mean		HE	-.01	.01	-.02	-.03
2 x Standard deviation			(.02)	(.04)	(.06)	(.09)

Note. "From" and "To" indicate the bottom and top limits of the 95% confidence interval. *Method of analysis: GL = Glass, McGaw, & Smith (1981); HU = Hunter, Schmidt, & Jackson (1982); HO = Hedges & Olkin (1985); HE = Hedges (1988).

CHAPTER V
CONCLUSIONS AND DISCUSSION

It was assumed in this dissertation that the teacher's task is to assist as many as possible of her or his students to pass an examination at the conclusion of the course (with as high a score as possible). The objective of this dissertation was to determine the correlation between teacher clarity of communication and the achievement gain of the students.

The population of students and teachers assumed to be covered by this study was all classes in public institutions (Grade 1 through undergraduate) where the education is of the American (European) type and the students or teachers are not selected as being in anyway exceptional.

Questions Answered in This Dissertation

The answers to the questions posed in this study were as follows:

1. What is the strength of the relationship between teacher clarity and student learning? The correlation between teacher clarity (corrected for unreliability in measurement) and mean class achievement gain (uncorrected for unreliability in measurement) is referred to as the effect size in the following. The effect size was .35 +/- .05.

2. Do clarity of (a) organization of the lesson (and course), (b) explanation (and speech), (c) examples and guided practice, and (d) assessment of student learning have different relationships to student learning? The effect sizes were (a) organization $.32 \pm .06$; (b) explanation $.35 \pm .08$ (speech $.43 \pm .14$); (c) examples and guided practice $.23 \pm .06$; and (d) assessment of student learning $.32 \pm .08$. These results overlap each other so they are not significantly different. In the regression equations the type of teacher clarity is not related to \underline{r} .

When two or more dimensions were rated at the same time (teacher skill--SKI), the effect size was $.60 \pm .13$ which is significantly higher than the single dimensions of teacher clarity (except for clarity of speech). Teacher behaviors were only classified as SKI at the college level (Grade 13) and the effect size is significantly related to grade, so the higher effect size might be due to this relationship.

3. Do student ratings of teacher clarity have a higher correlation with student learning than observer ratings? The effect size for student ratings was $.40 \pm .06$ and that for observer ratings $.28 \pm .05$. Student ratings do have a higher correlation than observer ratings, but student ratings tend to take place in college so the effect might be due to this relationship between rater and grade.

4. Is teacher clarity more important in subjects based on student verbal ability or in those based on numerical

ability? For both verbal and numerical subjects the effect size was $.35 \pm .05$.

5. Is teacher clarity more predictive of student learning at college, at secondary school, or at elementary school? Does the accuracy of prediction vary with grade? The effect sizes were (a) elementary school $.26 \pm .05$; (b) secondary school $.30 \pm .06$; and (c) college $.41 \pm .06$. The effect size for college was significantly higher than that for elementary school and was higher, but not significantly so, than that for secondary school. The correlation between effect size and grade (putting college as Grade 13) was about $.3$, and grade had a significant positive regression coefficient when effect size was modeled in terms of the variables in this study. The accuracy of prediction of teacher clarity does increase with grade level.

6. Is teacher clarity more predictive in large classes? With large classes (30 or more students) the effect size was $.42 \pm .11$ and with small classes $.28 \pm .05$. This is not a significant difference. The number of students in the class did not produce a significant regression coefficient, and the correlation with effect size was only $.13$. Thus it has not been established that teacher clarity is more predictive in large classes.

7. Does teacher clarity have a stronger relationship with student learning when the teacher is experienced than when she or he is inexperienced? The effect size for

experienced teachers was $.32 \pm .05$ and that for inexperienced teachers $.31 \pm .07$. The effect size was not different, but the coded variable for teacher experience did have a significant regression coefficient in the model. One can conclude that teacher clarity does have a stronger relationship with student learning for experienced teachers but that the effect is masked by the high negative correlation ($-.6$) between teacher experience and grade (many teaching assistants were studied at college level). The low values of effect size obtained here ($.31$ and $.32$) compared to the overall value of $.35$ are explained by the fact that the seven studies that included (and did not separate) the results for both experienced and inexperienced teachers produced an effect size of $.47 \pm .14$.

8. Which factors present in the investigation of relationships between teacher clarity and student learning are likely to result in an inaccurate estimation of the correlation? The studies that produced effect sizes more than two standard deviations from the mean did not apparently have anything in common.

Published studies produced an effect size of $.37 \pm .05$ compared to that for unpublished studies of $.21 \pm .05$. This is a significant difference but publication did not result in a significant regression coefficient, so a large part of this difference is explained by the correlation between publication and both grade and student ratings (both of which had correlations with effect size of about $.3$).

The studies using a large number of classes (40 or more) had an effect size of $.32 \pm .07$ and the others $.34 \pm .04$. The difference is not significant.

Experimental classes (less than 6 weeks--some only one lecture) produced the same effect size as normal classes.

When only a posttest was used and there was no evidence of random assignment of students to classes the effect size was $.60 \pm .09$. This was significantly higher than when a measure of achievement gain was used. One must conclude that posttest-only designs without random assignment are unsatisfactory in estimating the correlation between teacher clarity and class achievement gain.

When a simple-gain measure was used the effect size was $.33 \pm .07$ and with residual gain $.26 \pm .05$: The difference is not significant. The best design is to have random assignment of students to classes or to have the same students rate different teachers. In this case the effect size was $.41 \pm .16$. The confidence interval is so wide because there were only eight effect sizes.

9. Do the confidence intervals around the mean correlations obtained in these various circumstances vary significantly with the methods of analysis used? If they do, which method is likely to produce the most valid interval? If they do not, which is the easiest method? All methods gave practically the same results. It is not likely to be worth using an elaborate method like Tukey's jackknife. The easiest method was that of Glass et al. (1981).

Discussion

This study found that the correlation between any dimension of teacher clarity of communication and mean class achievement gain (effect size) was about $.35 \pm .05$. This indicates that teacher clarity is an important teacher characteristic.

The fact that there were no significant differences between the effect sizes of the dimensions of teacher clarity raises the issue of whether clarity of organization, clarity of explanation, examples and guided practice, and assessment of student learning are not separate dimensions and that teacher clarity is a unifactor quality. Against this view is the fact that when more than one of the dimensions were combined in a factor (teaching skill--SKI), the effect size increased to about $.6$. Was this because more than one teacher skill was being measured, or was it because teacher clarity was being measured more accurately? It seems possible that if all four dimensions were combined in a factor score, the effect size might be greater than $.6$.

Student ratings produce a higher effect size than do observer ratings. This is not surprising, as the students are in the class all the time and are better able to judge what is, or is not, clear to them.

It is sometimes suggested that teachers of subjects based on the students' numerical skills are at a disadvantage, compared to teachers of subjects based on the students' verbal skills, in obtaining high ratings on

student evaluations. This study did not address the relative size of the means in the two subject areas, but there was no difference in the effect sizes. Thus, teacher evaluations, at least for teacher clarity, are equally valid in both areas for predicting student learning.

Teacher clarity is more predictive of student learning at college than it is at school. Is it because the standard of teaching is more variable at college (the teachers are largely untrained), because there is less two-way communication at college than at school (often larger classes at undergraduate level), because more information is transferred by the lecture method, or for some other reason? This question bears investigation.

Teacher clarity is a better predictor of student learning for experienced teachers than it is for inexperienced teachers. Inexperienced teachers might have more basic problems (like controlling the class or lack of knowledge of the subject) that might be the cause of this.

Research using experimental classes was successful in obtaining the same correlation as in regular classes. Thus, investigations often involving the measurement of teacher clarity and student learning in a single lecture might be a valid means of investigating relationships likely to hold up in regular classes.

Posttest-only designs are unsatisfactory as a measure of student learning. There is no evidence that residual-gain measures produce a different effect size than

simple-gain measures. This is in agreement with the arguments of Rogosa and his colleagues (Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1985) and of Zimmerman and Williams (1982), that difference scores are reasonably reliable.

It is not likely to make any difference which of the three methods of meta-analysis one uses, so the analyst is free to go with her or his preferences.

One interesting relationship revealed in doing this study is that there is a strong correlation between ratings of teacher clarity and student learning when the reliability of the clarity rating is judged to be high (correlation between reliability and effect size is .66). There are only 10 studies that report both the reliability of the clarity measure and at least one effect size, so one is left to wonder whether this relationship is a stable one and, if so, what causes it. Are the studies better in some way when the reliability of the teacher clarity measure is high? The reliability of clarity measure was significantly related to 7 of the 10 other variables even though there were only 24 effect sizes from 10 studies. It would probably be productive to investigate some of these relationships.

This study has been successful in determining the confidence interval of the relationship between teacher clarity of communication and student learning in the class. An effect size of .35 indicates that if the average score of classes of similar students on a test is 50%, the students

in a class where the teacher is rated high on clarity of communication are likely to have a mean score in the region of 67% ($50 + \underline{r}/2$; Rosenthal & Rubin, 1982), whereas the average score of students in a class where the teacher is rated low on teacher clarity is likely to be in the region of 33% ($50 - \underline{r}/2$). Thus, an effect size of .35 indicates an important practical relationship.

The study has also related clarity of communication in the classroom to communication theory and has suggested possible dimensions of teacher clarity. The study might therefore contribute to the theory of teaching.

APPENDIX A
STUDIES USED IN THE META-ANALYSIS

Table A-1. Teaching Behaviors and Assumed Dimensions of Teacher Clarity

ID	Study	Teaching Behaviors	Dimension
1	Armento (1977)	Gives concept definition	EXP
		Gives concept example	EXP
		Asks for concept definition	ASL
		Asks for concept example	ASL
		Asks low-order question	ASL
		Asks high-order question	ASL
		Reviews, summarizes	ORG
2	Berliner & Tikunoff (1977)	T actively listens to S	ASL
		T checks S progress regularly	ASL
		T asks open-ended question	ASL
		T appears to perceive learning rate and adjusts teaching accordingly	EXP
		T seems confident teaching	EXP
		S copies T behavior	EGP
		T prepares S for lesson by reviewing, outlining, etc.	ORG

continued

Table A-1--continued

3	Brasskamp, Caulley, & Costin (1979)	TEACHER SKILL* 1. Put material across in an interesting way 2. Stimulated intellectual curiosity 3. Explained clearly and to the point 4. Skillful in observing student reactions 5. General (all-round) teaching ability	EXP
		TEACHER CONTROL* 1. Defined objectives of discussion 2. Controlled direction of discussion 3. Defined content of discussion 4. Asked specific, drill-type questions	ORG
4	Brophy & Evertson (1974)	S show clear understanding of T presentations	EXP
5	Bryson (1974)	Presents lecture material in an expressive and clear manner	EXP
		Carefully listens to and evaluates student opinion	ASL
		Gives clear and concise answers to questions	EGP
6	Doyle & Crichton (1978)	Clearly presented subject	EXP
7	Doyle & Whitely (1974)	Expositional skills	EXP
8	Crocker & Brooker (1986)	Teacher presentation	EXP

continued

Table A-1--continued

ID	Study	Teaching Behaviors	Dimension
9	Dunkin (1978)	Number of vague terms by T	SP
		T structuring (coverage)	ORG
		T structuring (repetition)	ORG
		Relevant high-level quest.	ASL
		Relevant low-level quest.	ASL
10	Centra (1977)	Course objectives & organization	ORG
11	Frey, Leonard, & Beatty (1975)	Presentation clarity	EXP
12	Good & Grouws (1979)	Conducts review	ORG
		Summarizes previous day's material	ORG
		Checks homework	ASL
		S accountable for seatwork	ASL
		S accountable for practice	ASL
		Demonstrations during presentation	EGP
		T-conducted seatwork (SW)	EGP
		T actively engages S in SW	EGP
T available for help in SW	EGP		
13	Hoffman	T KNOWLEDGE AND SKILL*	SKI
		1. Explained how the topics were related to each other	
		2. Used examples	
		3. Knew the subject matter	
		T asked S questions	ASL

continued

Table A-1--continued

ID	Study	Teaching Behaviors	Dimension
14	Marsh & Overall (1980)	ORGANIZATION* 1. Course materials and objectives were clearly outlined 2. Class presentations were well-prepared ENTHUSIASM/CONCERN* 1. T was enthusiastic 2. T gave presentations that made the subject understandable 3. T was concerned that S understood INTERACTION 1. S were encouraged to ask meaningful questions, to seek help, and to express their own ideas	ORG EXP ASL
15	Page (1958)	T-chosen comments on tests versus no comments	EGP
16	L. Smith (1985)	Absence of vagueness terms Lesson structure (kinetic)	SP EXP
18	Solomon, Rosenberg, & Bezdek (1964)	Clarity & Expressiveness vs Obscurity & Vagueness	EXP
19	Sullivan & Skanes (1974)	OVERALL RATING* 1. Presents material in a clear and easily understood manner 2. Gets Ss really interested in the subject 3. Interest in students	EXP

continued

Table A-1--continued

ID	Study	Teaching Behaviors	Dimension
17	Orpen (1980)	SKILL*	EXP
		1. All-round teaching ability	
		2. T ability in observing S reactions	
		3. Stimulating the intellectual curiosity of the S	
		4. Explaining clearly and to the point	
		5. Puts material across in an interesting way	
		STRUCTURE*	ORG
		1. Deciding what should be done and how	
		2. Following an outline closely	
		3. Concern for keeping a tight schedule	
		FEEDBACK*	EGP
		1. Telling the S when they have done a particularly good job	
2. Complimenting S in front of others			
3. Criticizing poor work			
4. Keeping S well-informed of their progress			
20	Wright & Nuthall (1970)	Terminal structuring	ORG
		Recapitulation	ORG
		Review	ORG
		Structuring total	ORG
21	Gage, Belgard, Dell, Hiller, Rosenshine, & Unruh (19680)	Presentation clarity	EXP
		Pacing the lecture	EXP
		Clarity of aims	ORG
		Organization of lecture	ORG

continued

Table A-1--continued

ID	Study	Teaching Behaviors	Dimension
22	Flanders (1970)	T questions	ASL
23	Benton (1976)	CONTENT MEANINGFUL* 1. T lectures are not over my head 2. T speaks clearly 3. T makes the connection clear between ideas 4. T explains clearly PLANNING & LEARNING CLIMATE* 1. T used classtime well 2. T accomplished objectives 3. Objectives for the course were made clear 4. T summarized or emphasized major points.	EXP ORG
24	Morsh, Burgess, & Smith (1956)	(Reliability measure only)	
25	Peterson, Micceri, & Smith (1985)	(Reliability measure only)	
26	Poonyakanok, Thisayakorn, & Digby (1986)	(Reliability measure only)	
27	Bourke (1985)	Extent of coverage of posttest items Review to refresh learning Homework used T gave help during lessons Total number of T questions	ORG ORG EGP EGP ASL
29	Costin (1978)	T SKILL* 1. T put material across in an interesting way 2. T stimulated the intellectual curiosity of the S 3. T was skillful in observing S reactions 4. Overall rating of T	SKI

continued

Table A-1--continued

ID	Study	Teaching Behaviors	Dimension
28	Costin, Greenough, & Menges (1971)	(Reliability measure only)	
30	Evertson & Brophy (1974)	S show clear understanding of T presentations	EXP
		T goes to seats to check work	ASL
		T uses advance organizers to introduce activities	ORG
		T well-organized, prepared	ORG
		T goes to Ss desk to give help	EGP
31	Hiller, Fisher, & Kaess (1969)	Lack of vagueness	SP
32	Hines, Cruickshank, & Kennedy (1985)	TEACHER CLARITY* 1. T stresses important aspects of content 2. T explains content by use of examples 3. T assesses and responds to perceived deficiencies in understanding	SKI
33	Leclerc, Bertrand, & Dufour (1986)	T asks low-level question	ASL
34	Lorentz (1977)	Ability to communicate effectively with S	EXP
		Pauses, elicits, and responds to S questions	EGP
		Utilizes S feedback to modify teaching	ASL
35	McDonald & Elias (1976)	Directed S seatwork	EGP
36	Marsh (1987)	(Reliability measures only)	

continued

Table A-1--continued

ID	Study	Teaching Behaviors	Dimension
37	Pinney (1970)	T announcements about important points	EXP
		Vocal intensity: significant variation in pitch and/or volume	SP
38	Shavelson & Dempsey-Atwood (1976)	(Reliability measures only)	
39	Austin (1976)	Comments on homework versus just grading	EGP
40	McKeachie, Linn, & Mann (1971)	SKILL* 1. All-round teaching ability 2. T ability in observing S reactions 3. Stimulating the intellectual curiosity of the S 4. Explaining clearly and to the point	EXP
		STRUCTURE* 1. T decided in detail what should be done and how and how it should be done 2. T followed an outline closely 3. T had everything going according to schedule	ORG
		FEEDBACK* 1. T told S when they had done a particularly good job 2. T complimented S in front of others 3. T criticized poor work	EGP
41	Marsh, Fleiner, & Thomas (1975)	Class presentation	EXP

continued

Table A-1--continued

ID	Study	Teaching Behaviors	Dimension
42	Hazelton (1980)	FACILITATION OF LEARNING* 1. S could understand class presentations 2. T delivered orderly, logical presentations of material 3. T spoke with poise 4. T gave organized answers to complicated questions in class	EXP
43	Gessner (1973)	Content and organization Presentation	ORG EXP
44	Frey (1973)	Planning and organization Presentation Class discussion	ORG EXP ASL
46	Ellis & Rickard (1977)	(Reliability measures only)	
47	Foy (1969)	(Reliability measures only)	

Note. T = teacher, S = student.

* A factor defined by the numbered behaviors

Table A-2. Characteristics and Results of Studies

ID	VER	PUB	STU	ACH	REL	TEX	WKS	GRA	NS	NC	DIM	R	SMR	SMZ
1	1	1	-1	2	.85	-1	.4	5	14	22	EXP	.46	.	.
1	1	1	-1	2	.85	-1	.4	5	14	22	ASL	.06	.	.
1	1	1	-1	2	.85	-1	.4	5	14	22	ORG	.38	.30	.31
2	-1	1	-1	.	.	1	30	2	30	20	ASL	.40	.	.
2	-1	1	-1	.	.	1	30	2	30	20	EXP	.45	.	.
2	-1	1	-1	.	.	1	30	2	30	20	EGP	.32	.	.
2	-1	1	-1	.	.	1	30	2	30	20	ORG	.32	.	.
2	1	1	-1	.	.	1	30	2	30	20	ASL	.49	.	.
2	1	1	-1	.	.	1	30	2	30	20	EXP	.53	.	.
2	1	1	-1	.	.	1	30	2	30	20	EGP	.27	.	.
2	1	1	-1	.	.	1	30	2	30	20	ORG	.50	.	.
2	-1	1	-1	.	.	1	30	5	30	20	ASL	.34	.	.
2	-1	1	-1	.	.	1	30	5	30	20	EXP	.32	.	.
2	-1	1	-1	.	.	1	30	5	30	20	EGP	.06	.	.
2	-1	1	-1	.	.	1	30	5	30	20	ORG	.42	.	.
2	1	1	-1	.	.	1	30	5	30	20	ASL	.33	.	.
2	1	1	-1	.	.	1	30	5	30	20	EXP	.44	.	.
2	1	1	-1	.	.	1	30	5	30	20	EGP	.17	.	.
2	1	1	-1	.	.	1	30	5	30	20	ORG	.37	.38	.40
3	1	1	1	1	.85	-1	8	13	25	17	EXP	.30	.	.
3	1	1	1	1	.85	-1	8	13	25	17	ORG	.33	.32	.33
4	-1	-1	-1	3	.	1	30	2.5	30	30	EXP	.07	.	.
4	1	-1	-1	3	.	1	30	2.5	30	30	EXP	.15	.11	.11
5	-1	1	1	.	.	1	30	2	30	20	EXP	.37	.	.
5	-1	1	1	.	.	1	30	2	30	20	ASL	.39	.	.
5	-1	1	1	.	.	1	30	2	30	20	EGP	.34	.37	.41
6	1	1	1	3	.78	-1	15	13	21	12	EXP	.20	.20	.20
7	1	1	1	3	.	-1	15	13	15	12	EXP	.31	.31	.32
8	-1	1	-1	2	.	1	30	2	30	36	EXP	.15	.	.
8	1	1	-1	2	.	1	30	2	30	36	EXP	.36	.	.
8	-1	1	-1	2	.	1	30	5	30	36	EXP	.18	.	.
8	1	1	-1	2	.	1	30	5	30	36	EXP	.15	.12	.12
9	1	1	-1	3	.	1	.2	6	35	29	SP	.08	.	.
9	1	1	-1	3	.	1	.2	6	35	29	ORG	.30	.	.
9	1	1	-1	3	.	1	.2	6	35	29	ASL	.10	.16	.16
10	-1	1	1	1	.	1	15	13	.	30	ORG	.30	.	.
10	1	1	1	1	.	1	15	13	.	30	ORG	.48	.39	.41

continued

Table A-2--continued

ID	VER	PUB	STU	ACH	REL	TEX	WKS	GRA	NS	NC	DIM	<u>R</u>	<u>SMR</u>	<u>SMZ</u>
11	-1	1	1	2	.	-1	15	13	26	17	EXP	.42	.	.
11	-1	1	1	2	.	-1	15	13	26	17	ASL	.21	.	.
11	1	1	1	2	.	-1	15	13	26	17	EXP	.81	.	.
11	1	1	1	2	.	-1	15	13	26	17	ASL	.40	.46	.50
12	-1	1	-1	3	.	1	30	4	25	40	ORG	.29	.	.
12	-1	1	-1	3	.	1	30	4	25	40	ASL	.34	.	.
12	-1	1	-1	3	.	1	30	4	25	40	EGP	.18	.27	.28
13	1	1	1	1	.82	0	15	13	30	142	SKI	.56	.	.
13	1	1	1	1	.82	0	15	13	30	142	ASL	.59	.	.
13	1	1	1	3	.82	0	15	13	30	142	SKI	.29	.	.
13	1	1	1	3	.82	0	15	13	30	142	ASL	.36	.45	.49
14	-1	1	1	2	.	-1	15	13	31	31	ORG	.22	.	.
14	-1	1	1	2	.	-1	15	13	31	31	EXP	.40	.	.
14	-1	1	1	2	.	-1	15	13	31	31	ASL	.36	.33	.34
15	0	1	-1	3	.	1	3	9	9	149	EGP	.14	.14	.14
16	-1	1	-1	3	.91	1	.1	10	25	19	SP	.47	.	.
16	-1	1	-1	3	.91	1	.1	10	25	19	EXP	.51	.49	.54
17	-1	1	1	3	.	-1	15	13	10	10	ORG	.10	.	.
17	-1	1	1	3	.	-1	15	13	10	10	SKI	.61	.	.
17	-1	1	1	3	.	-1	15	13	10	10	EGP	.21	.31	.32
18	1	1	-1	1	.	-1	15	13	25	24	EXP	.58	.58	.66
19	-1	1	1	4	.	0	15	13	25	70	EXP	.41	.	.
19	1	1	1	4	.	0	15	13	25	70	EXP	.21	.	.
19	-1	1	1	4	.	1	15	13	25	70	EXP	.53	.	.
19	-1	1	1	4	.	-1	15	13	25	70	EXP	.01	.29	.30
20	-1	1	-1	3	.	0	.6	3	25	17	ORG	.07	.07	.07
21	1	-1	1	4	.	1	.1	12	21	43	EXP	.26	.	.
21	1	-1	1	4	.	1	.1	12	21	43	ORG	.32	.29	.29
22	0	1	-1	3	.	1	2	2	25	15	ASL	.07	.	.
22	1	1	-1	3	.	1	2	4	25	16	ASL	-.19	.	.
22	0	1	-1	3	.	1	12	6	25	30	ASL	.11	.	.
22	1	1	-1	3	.	1	2	7	25	15	ASL	-.06	.	.
22	-0	1	-1	3	.	1	2	8	25	16	ASL	.44	.10	.10
23	1	-1	1	0	.	-1	10	13	17	31	EXP	-.19	.	.
23	1	-1	1	0	.	-1	10	13	17	31	ORG	.11	-.04	-.04
24	-1	1	1	3	.42	1	2	12	14	102

continued

Table A-2--continued

ID	VER	PUB	STU	ACH	REL	TEX	WKS	GRA	NS	NC	DIM	R	SMP	SMZ
2586
2671
27	-1	1	-1	.	.	1	12	10	25	75	ORG	.30	.	.
27	-1	1	-1	.	.	1	12	10	25	75	EGP	.24	.	.
27	-1	1	-1	.	.	1	12	10	25	75	ASL	.00	.18	.18
2883
29	1	1	1	4	.	-1	16	13	14	96	SKI	.49	.49	.54
30	1	-1	-1	3	.75	1	30	2.5	25	75	EXP	.20	.	.
30	1	-1	-1	3	.75	1	30	2.5	25	75	ASL	.02	.	.
30	1	-1	-1	3	.75	1	30	2.5	25	75	ORG	.27	.	.
30	1	-1	-1	3	.75	1	30	2.5	25	75	EGP	.18	.	.
30	-1	-1	-1	3	.75	1	30	2.5	25	75	EXP	.17	.	.
30	-1	-1	-1	3	.75	1	30	2.5	25	75	ASL	-.14	.	.
30	-1	-1	-1	3	.75	1	30	2.5	25	75	ORG	.04	.	.
30	-1	-1	-1	3	.75	1	30	2.5	25	75	EGP	.19	.12	.12
31	1	1	-1	3	.	1	.1	12	21	55	SP	.54	.54	.60
32	-1	1	-1	1	.86	-1	.2	13	5	32	SKI	.68	.	.
32	-1	1	1	1	.87	-1	.2	13	5	32	SKI	.61	.65	.78
33	-1	1	-1	3	.	1	12	7.5	30	29	ASL	.11	.11	.11
34	0	-1	-1	3	.45	1	20	5.5	30	26	EXP	-.03	.	.
34	0	-1	-1	3	.45	1	20	5.5	30	26	ASL	-.73	-.19	-.19
35	1	-1	-1	2.5	.45	1	30	2	20	41	EGP	.4	.	.
35	1	-1	-1	2.5	.45	1	30	5	20	41	EGP	.4	.	.
35	-1	-1	-1	2.5	.45	1	30	2	20	41	EGP	.1	.	.
35	-1	-1	-1	2.5	.45	1	30	5	20	41	EGP	-.2	.18	.18
36	0	1	1	.	.77	.	15	13	100
37	1	-1	-1	3	.	-1	.2	8.5	25	32	EXP	.35	.	.
37	1	-1	-1	3	.	-1	.2	8.5	25	32	SP	.35	.35	.37
38	0	1	-1	.	.63
39	-1	1	-1	3	.	1	6	7	12	18	EGP	.47	.47	.51
40	1	1	1	3	.	-1	15	13	25	143	EXP	.14	.	.
40	1	1	1	3	.	-1	15	13	25	143	ORG	.01	.	.
40	1	1	1	3	.	-1	15	13	25	143	EGP	.14	.10	.10

continued

Table A-2--continued

ID	VER	PUB	STU	ACH	REL	TEX	WKS	GRA	NS	NC	DIM	<u>R</u>	<u>SMR</u>	<u>SMZ</u>
41	-1	1	1	3	.	-1	15	13	40	18	EXP	.43	.43	.46
42	-1	-1	1	3	.89	-1	10	13	30	36	EXP	-.25	-.25	.26
43	-1	1	1	4	.	1	15	13	78	20	ORG	.77	.	.
43	-1	1	1	4	.	1	15	13	78	20	EXP	.69	.73	.93
44	-1	1	1	3	.	1	10	13	35	13	ORG	.67	.	.
44	-1	1	1	3	.	1	10	13	35	13	EXP	.79	.73	.93
45	-1	1	1	3	.	1	15	13	35	53	ORG	.68	.	.
45	-1	1	1	3	.	1	15	13	35	53	EXP	.62	.	.
45	-1	1	1	3	.	1	15	13	35	53	ASL	.37	.56	.63
4687
4771

Note. ID: identification number of study; VER: 1 = subject based on students' verbal ability, -1 = subject based on students' numerical ability, 0 = both subject areas; PUB: 1 = study published in a journal or book, -1 = dissertation or ERIC; STU: 1 = student rating of teacher clarity, -1 = observer rating of teacher clarity, 0 = rating by both students and observers; ACH: 0 = essay test rated by the class teacher, 1 = posttest only with no random assignment of students, 2 = simple gain score, 3 = residual gain score, 4 = random assignment of students or the same students rate all the teachers; REL = reliability of the teacher-clarity measure; TEX: 1 = experienced teachers, -1 = student teachers or teaching assistants; WKS = weeks between the start of the course and the posttest; GRA = grade (college = 13, 8.5 = grades 8 and 9); NS = average number of students in a class; NC = number of classes in the study; DIM: ORG = clarity of organization, EXP = clarity of explanation, EGP = examples and guided practice, ASL = assessment of student learning, SP = clarity of speech, SKI = a factor score comprising more than one dimension and no one dimension is dominating the factor; R = the correlation between the dimension of teacher clarity and the achievement gain of the class (the average value of all the rs reported in the study for the dimension, the grade, and the subject area); SMR = the study mean value of R averaged over all the Rs for the study; SMZ = the Fisher z equivalent of SMR.

APPENDIX B
REJECTED STUDIES

Table B-1. Rejected Studies and Reason for Rejection

Study	Reason for Rejection
Aagard (1973)	No class-level <u>rs</u>
Abrami & Mizener (1982)	Grades used as achievement measure
Amidon & Giammatteo (1967)	Superior T nominated by administrators
Aubrect (1981)	Review
Baird (1983)	No data
Bendig (1953)	Overall teacher rating only
Bennett & Jordan (1976)	No class-level <u>rs</u>
Benton (1982)	Review
Beseda (1973)	No class-level <u>rs</u>
Blaney (1983)	No class-level <u>rs</u>
Brown (1977)	No TC variables
Bush, Kennedy, & Cruickshank (1977)	No achievement measures
Chase & Keene (1979)	No class-level <u>rs</u>
Clark, Gage, Marx, Peterson, Stayrook, & Winne (1979)	No class-level <u>rs</u>
Cobb (1972)	No T behaviors observed
Cook (1967)	No TC variables
Cooley & Leinhardt (1980)	No TC variables

continued

Table B-1--continued

Study	Reason for Rejection
Crawford, Evertson, Anderson, & Brophy (1976)	No TC variables
Creamer & Lorentz (1979)	No class-level <u>rs</u>
Cruickshank (1985)	Applications of TC only
Domino (1971)	No TC variables
Doyle (1979)	No TC variables
Druva & Anderson (1983)	T Characteristics not behaviors
Duffy, Roehler, Meloth, & Vaurus (1986)	Qualitative research
Ellis & Rickard (1977)	Overall rating of T only
Endo & Della-Piano (1976)	No class-level <u>rs</u>
Evans & Guyman (1978)	No class-level <u>rs</u>
Evertson, Anderson, Anderson, & Brophy (1980)	No class-level <u>rs</u>
Evertson, Anderson, & Brophy (1979)	No class-level <u>rs</u>
Feldman (1976)	Review
Follman (1983)	Review
Good & Brophy (1974)	No TC variables
Good & Grouws (1977)	No class-level <u>rs</u>
Goodlad & Klein (1974)	No class-level <u>rs</u>
Green (1983)	No TC variables
Hammer (1972)	No class-level <u>rs</u>
Heil, Powell, & Felfer (1960)	No TC variables
Hsu & White (1978)	Canonical correlations reported between achievement and unknown factors of student assessment

continued

Table B-1--continued

Study	Reason for Rejection
Land (1979)	No class-level <u>rs</u>
Land (1980)	No class-level <u>rs</u>
Land (1981a)	No class-level <u>rs</u>
Land (1981b)	No class-level <u>rs</u>
Land & Denham (1979)	No class-level <u>rs</u>
Lorentz & Coker (1979)	No class-level <u>rs</u>
Marsh (1977)	No class-level <u>rs</u>
Marsh (1982)	No achievement measure
Martikean (1973)	No class-level <u>rs</u>
Mathis & Shrum (1977)	No class-level <u>rs</u>
Medley & Mitzell (1959)	No achievement measures
McKeachie & Kulik, (1975)	Review article
McKeachie & Linn (1978)	No TC variables, no achievement
McKeachie, Linn, & Mendelson (1978)	Overall rating of T only
McKeachie & Solomon (1958)	No achievement measure
McKinney, Mason, Parkinson, & Clifford (1975)	Only S behavior reported
Morsh, Burgess, & Smith (1956)	Overall rating of T only
Peterson, Micceri, & Smith (1985)	No TC variables
Peterson (1979)	No TC variables
Peterson, Marx, & Clark (1978)	No TC variables
Pitman (1985)	No achievement measures

continued

Table B-1--continued

Study	Reason for Rejection
Rodin & Rodin (1972)	Overall rating of T only
Rosenshine (1970a)	Review
Rosenshine (1970b)	Review
Ryan (1973)	No class-level <u>rs</u> . Controls were not taught relevant content
Ryan (1974)	No class-level <u>rs</u>
Savage (1972)	No class-level <u>rs</u>
Sharp (1966)	No TC variables
L. Smith (1979)	No class-level <u>rs</u>
L. Smith (1985b)	No class-level <u>rs</u>
L. Smith (1985c)	No class-level <u>rs</u>
Smith & Cotton (1980)	No class-level <u>rs</u>
Smith & Sanders (1981)	No class-level <u>rs</u>
A. Snider (1965)	No TC variables
R. Snider (1966)	No class-level <u>rs</u>
Soar (1968)	No TC variables
Soar (1972)	Review
Soar (1973)	No TC variables
Soar & Soar (1973)	No TC variables
Solomon & Kendall (1976)	No TC variables
Stallings (1974)	No TC variables
Stallings (1977)	No TC variables
Stallings & Kaskowitz (1974)	No TC variables
Tobin & Capie (1982)	No class-level <u>rs</u>
Torrance & Parent (1966)	No class-level <u>rs</u>

continued

Table B-1--continued

Study	Reason for Rejection
Trincherro (1974)	No TC variables
Trincherro (1975)	No TC variables
Trindade (1972)	No class-level <u>rs</u>
Turner & Thompson (1973)	The first exam after 3 weeks (which correlated .73 with TC) was used as the pretest. Thus TC is largely partialled out of the <u>rs</u> .
Vorrayer (1969)	No TC variables
Zelby (1974)	No class-level <u>rs</u>

Note. T = teacher, S = student, TC = teacher clarity.

REFERENCES

- Aagard, S. A. (1973). Oral questioning by the teacher: Influence on student achievement in eleventh grade chemistry. Dissertation Abstracts International, 34, 631A. (University Microfilms No. 73-19406)
- Abrami, P. C., & Mizener, D. A. (1982, August). Student/instructor attitude similarity, course ratings, and student achievement. Paper presented at the Meeting of the American Psychological Association, Washington, DC. (ERIC Document Reproduction Service No. ED 223 144)
- Amidon, E. J., & Giammatteo, M. (1967). The verbal behavior of superior elementary school teachers. In E. J. Amidon & J. B. Hough (Eds.), Interaction analysis: Theory, research and application (pp. 186-187). Reading, MA: Addison-Wesley.
- Armento, B. (1977). Teacher behaviors related to student achievement on a social science concept test. Journal of Teacher Education, 28(2), 46-52.
- Aubrect, J. D. (1981). Reliability, validity, and generalizability of student ratings of instruction (IDFA Paper No. 6). Kansas State University: Center for Faculty Evaluations and Development. (ERIC Document Reproduction Service No. ED 213 296)
- Austin, J. D. (1976). Do comments on mathematics homework affect student achievement? School Science and Mathematics, 76, 159-164.
- Baird, J. S. (1983). Validity and reliability of student ratings of faculty. Teaching of Psychology, 10(1), 46.
- Bendig, A. W. (1953). The relation of level of course achievement to students' instructor and course rating in introductory psychology. Educational and Psychological Measurement, 13, 437-448.
- Bennett, N., & Jordan, J. (Eds.) (1976). Teaching styles and pupil progress. Cambridge, MA: Harvard University Press.

- Benton, S. E. (1976, April). A comparison of the criterion validity of two types of student response inventories for appraising instruction. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco. (ERIC Document Reproduction Service No. ED 128 397)
- Benton, S. E. (1982). Rating college teaching: Criterion validity studies of student evaluation of instruction instruments. AAHE-ERIC/Higher Education Research Report No. 1. (ERIC Document Reproduction Service No. ED 221 147)
- Berliner, D. C., & Tikunoff, W. (1977). Ethnography in the classroom. In G. Borich (Ed.), The appraisal of teaching: Concepts and processes (pp. 280-291). Reading, MA: Addison-Wesley.
- Beseda, C. G. (1973). Levels of questioning used by student teachers and the effect on pupil achievement and critical thinking ability. Dissertation Abstracts International, 33, 4214A. (University Microfilms No. 73-2889)
- Bourke, S. F. (1985). The study of classroom contexts and practices. Teaching and Teacher Education, 1, 33-50.
- Blaney, R. (1983). Effects of teacher structuring and reacting on student achievement. Elementary School Journal, 83, 569-577.
- Brasskamp, L. A., Caulley, D., & Costin, F. (1979). Student ratings and instructor self-ratings and their relationship to student achievement. American Educational Research Journal, 16, 295-306.
- Brophy, J. E. (1987). Teacher effects research and teacher quality. Journal of Classroom Interaction, 22(1), 14-23.
- Brophy, J. E., & Evertson C. M. (1974). Process-Product Correlations in the Texas Teacher Effectiveness Study (Final Report). Austin: University of Texas, Research & Development Center. (ERIC Document Reproduction Service No. ED 091 394)
- Brown, R. (1977). The relationship between student evaluation of teaching, student achievement, and student perception of teacher effectiveness. (ERIC Document Reproduction Service No. ED 133 314)
- Bryson, R. (1974). Teacher evaluations and student learning: A re-examination. The Journal of Educational Research, 68, 12-14.

- Bullock, R. J., & Svyantek, D. J. (1985). Analyzing meta-analysis: Potential problems, an unsuccessful replication, and evaluation criteria. Journal of Applied Psychology, 70, 108-115.
- Bush, A. J., Kennedy, J. J., & Cruickshank, D. R. (1977). An empirical investigation of teacher clarity. Journal of Teacher Education, 28(2), 53-58.
- Carlberg, C. G., & Kavale, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. The Journal of Special Education, 14, 295-309.
- Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. American Educational Research Journal, 14, 17-24.
- Centra, J. A. (1980). Determining faculty effectiveness. San Francisco: Jossey-Bass.
- Chase, C. I., & Keene, J. M. (1979). Validity of student ratings of faculty. Bloomington, IN: Indiana University, Bureau of Educational Studies and Testing. (ERIC Document Reproduction Service No. ED 169 870)
- Chilcoat, G. (1987). Teacher talk: Keep it clear. Academic Therapy, 22, 263-271.
- Clark, C. M., Gage, N. L., Marx, R. W., Peterson, P. L., Stayrook, N. G., & Winne, P. H. (1979). A factorial experiment on teacher structuring, soliciting and reacting. Journal of Educational Psychology, 71, 534-552.
- Cobb, J. A. (1972). Relationship of discrete classroom behaviors to fourth grade academic achievement. Journal of Educational Psychology, 63, 74-80.
- Cook, R. E. (1967). The effect of teacher methodology upon certain achievements of students in secondary school biology. Dissertation Abstracts International, 28, 3066A. (University Microfilms No. 68-912)
- Cooley, W., & Leinhardt, G. (1980). The instructional dimensions study. Educational Evaluation and Policy Analysis, 2(1), 7-25.
- Cooper, B., & Foy, J. M. (1967). Evaluating the effectiveness of lectures. Universities Quarterly, 21, 182-185.

- Costin, F. (1978). Do student ratings of college teachers predict student achievement? Teaching of Psychology, 5, 86-88.
- Costin, F., Greenough, W. T., & Menges, R. J. (1971). Student ratings of college teaching, reliability, validity, and usefulness. Review of Educational Research, 41, 511-535.
- Crawford, W. J., Evertson, C. M., Anderson, L. M. & Brophy, J. E. (1976). Process-product relationships in second and third grade classrooms (Report No 76-11). Austin: Texas University R. & D. Center for Teacher Ed. (ERIC Document Reproduction Service No. ED 148 888)
- Creamer, M., & Lorentz, J. L. (1979, November). Effect of teacher structure, teacher affect, cognitive level of questions, group size, and student social status on reading achievement. Paper presented at the meeting of the National Reading Conference, San Antonio, TX. (ERIC Document Reproduction Service No. ED 185 517)
- Crocker, R. K., & Brooker, G. M. (1986). Classroom control and student outcomes in grades 2 and 5. American Educational Research Journal, 23, 1-11.
- Cruickshank, D. R. (1985). Applying research on teacher clarity. Journal of Teacher Education, 36(2), 44-48.
- Cruickshank, D. R., & Kennedy, J. J. (1986). Teacher clarity. Teaching and Teacher Education, 2, 43-67.
- Cruickshank, D. R., Myers, B. & Moenjck, T. (1975). Statements of clear teacher behaviors provided by 1009 students in grades 6-9. Unpublished manuscript, The Ohio State University.
- Doyle, K. O., Jr., & Crichton, L. I. (1978). Student, peer, and self evaluations of college instructors. Journal of Educational Psychology, 70, 815-826.
- Doyle, K. O., Jr., & Whitely, S. E. (1974). Student ratings as criteria for effective teaching. American Educational Research Journal, 11, 259-274.
- Doyle, W. (1979). The tasks of teaching and learning in classrooms: Correlates of effective teaching (R. & D. Report No. 4103). Austin: Texas University R. & D. Center for Teacher Ed. Paper presented at the meeting of the A.E.R.E. (ERIC Document Reproduction Service No. ED 185 069)

- Druva, C. A., & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior by student outcome: A meta-analysis of research. Journal of Research in Science Teaching, 20, 467-479
- Duffy, G. G., Roehler, L. R., Meloth, M. S., & Vaurus, L. G. (1986). Conceptualizing instructional explanation. Teaching and Teacher Education, 2, 197-214.
- Ellis, N. R., & Rickard, H. C. (1977). Evaluating the teaching of introductory psychology. Teaching of Psychology, 4, 128-132.
- Endo, G. T., & Della-Piano, G. (1976). A validation study of course evaluation ratings. Improving College and University Teaching, 24, 84-86.
- Evans, W. E., & Guyman, R. E. (1978). Clarity of explanation: A powerful indicator of teacher effectiveness. Paper presented at the meeting of the American Educational Research Association, Toronto. (ERIC Document Reproduction Service No. ED 151 321)
- Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. American Educational Research Journal, 17, 43-60.
- Evertson, C. M., Anderson, L. M., & Brophy, J. E. (1979). Correlates of effective teaching. Texas Junior High School Study, Final Report of Process outcome relationships, Vol. 1. Austin: R. & D. Center for Teacher Education. (ERIC Document Reproduction Service No. ED 173 744)
- Evertson, C. M., & Brophy, J. E. (1974). High inference behavioral ratings as correlates of teaching effectiveness. Austin, TX: R. & D. Center for Teacher Education. (ERIC Document Reproduction Service No. ED 095 174)
- Feldman, K. A. (1976). The superior college teacher from the student's point of view. Research in Higher Education, 5, 243-288.
- Flanders, N. A. (1970). Analyzing teaching behavior. Reading, MA: Addison-Wesley.
- Follman, J. (1983, March). Student ratings of faculty teaching effectiveness: Revisited. Paper presented at the meeting of the Association for the Study of Higher Education, Washington, DC. (ERIC Document Reproduction Service No. ED 232 556)

- Foy, J. M. (1969). A note on lecturer evaluation by students. Universities Quarterly, 23, 345-349.
- Frey, P. W. (1973). Student ratings of teaching: Validity of several rating factors. Science, 182, 83-85.
- Frey, P. W. (1976). Validity of student instructional ratings: Does timing matter? Journal of Higher Education, 47, 327-336.
- Frey, P. W., Leonard, D. W., & Beatty, W. W. (1975). Student ratings of instruction: Validation research. American Educational Research Journal, 12, 435-447.
- Gage, N. L. (1979). The generality of dimensions of teaching. In P. L. Peterson & H. J. Walberg (Eds.), Research on teaching (pp. 264-288). Berkeley, CA: McCutchan.
- Gage, N. L., Belgard, M., Dell, D., Hiller, J. E., Rosenshine, B., & Unruh, W. R. (1968). Explorations of teachers' effectiveness in explaining (Tech. Rep. No. 4.). Stanford University, Stanford School of Education. (ERIC Document Reproduction Service No. ED 028 147)
- Gage, N. L., & Needels, M. C. (1989). Process-product research on teaching: A review of criticisms. Elementary School Journal, 89, 273-300.
- Gessner, P. K. (1973). Evaluation of Instruction. Science, 180, 566-570.
- Glass, G. V., Cahen, L., Smith, M. L., & Filby, N. (1983). School class size. Beverly Hills, CA: Sage.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.
- Good, T. L., & Grouws, D. A. (1977). Teaching effectiveness in fourth-grade mathematics classrooms. In G. Borich (Ed.), The appraisal of teaching: Concepts and processes (pp. 121-129). Reading, MA: Addison-Wesley.
- Good, T. L., & Grouws, D. A. (1979). Missouri mathematics effectiveness project. Journal of Educational Psychology, 71, 355-362.
- Good, T. L., Grouws, D. A., & Ebmeier, M. (1983). Active mathematics teaching. New York: Longman.

- Goodlad, J. I., & Klein, M. F. (1974). Looking behind the classroom door. Worthington, OH: C. A. Jones.
- Green, J. L. (1983). Research on teaching as a linguistic process: A state of the art. In E. W. Gordon (Ed.), Review of research in education (Vol. 10, pp. 151-252). Washington, DC: American Educational Research Association.
- Hammer, B. (1972). Grade expectations, differential teacher comments, and student performance. Journal of Educational Psychology, 63, 454-458.
- Hazelton, A. E. (1980). A study of the validity of student ratings of college teaching assessed on a criterion of student achievement in a first course in calculus (Doctoral Dissertation, University of Florida, 1980). Dissertation Abstracts International, 41, 3901A.
- Hedges, L. V. (1988). The meta-analysis of test validity studies: Some new approaches. In H. Weiner and H. I. Braun (Eds.), Test validity (pp. 191 - 212). Hillsdale, NJ: Lawrence Erlbaum.
- Hedges, L. V., & Olkin, I. (1984). Nonparametric estimators of effect size in meta-analysis. Psychological Bulletin, 96, 573-580.
- Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.
- Heil, L. M., Powell, M., & Felfer, I. (1960). Characteristics of teacher behavior related to the achievement of children in several elementary grades (Final Report). US Office of Education, Cooperative Research Branch (Research Project No. SAE 7285). (ERIC Document Reproduction Service No. ED 002 843)
- Hiller, J. H., Fisher, G. A., & Kaess, W. (1969). A computer investigation of verbal characteristics of effective classroom learning. American Educational Research Journal, 6, 661-665.
- Hines, C. V. (1981). A further investigation of teacher clarity: The observation of teacher clarity and the relationship between clarity and student achievement and satisfaction. Dissertation Abstracts International, 42, 3122A. (University Microfilms No. 8129015)
- Hines, C. V., Cruickshank, D. R., & Kennedy, J. J. (1985). Teacher clarity and its relationship to student achievement and satisfaction. American Educational Research Journal, 22, 87-99.

- Hoffman, R. G. (1978). Variables affecting university student ratings of instructor behavior. American Educational Research Journal, 15, 287-299.
- Hsu, V-M, & White, W. F. (1978, March). Interactions between teaching performance and student achievement. Paper presented at the meeting of the American Educational Research Association, Toronto, Canada. (ERIC Document Reproduction Service No. ED 151 332)
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.
- Johnson, D. W., Johnson, R., & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. Review of Educational Research, 53, 5-54.
- Kennedy, J. J., Cruickshank, D. R., Bush, A. J., & Myers, B. (1978). Additional investigations into the nature of teacher clarity. Journal of Educational Research, 72, 3-10.
- Kounin, J. S. (1970). Discipline and group management in classrooms. New York: Holt, Rinehart, & Winston.
- Kraemer, H. C., & Andrews, G. (1982). A nonparametric technique for meta-analysis effect size calculation. Psychological Bulletin, 91, 404-412.
- Kulik, C., & Kulik, J. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. American Educational Research Journal, 19, 415-428.
- Kurosawa, K. (1984). Meta-analysis and selective publication bias. American Psychologist, 39, 73-75.
- Land, M. L. (1979). Low-inference variables of teacher clarity: Effect on student concept learning. Journal of Educational Psychology, 71, 795-799.
- Land, M. L. (1980). Teacher clarity and cognitive level of questions: Effects on learning. Journal of Experimental Education, 49, 48-51.
- Land, M. L. (1981a). Actual and perceived teacher clarity--relations to student achievement in science. Journal of Research in Science Teaching, 18, 139-143.

- Land, M. L. (1981b). Combined effect of two teacher clarity variables on student achievement. Journal of Experimental Education, 50, 14-17.
- Land, M. L., & Denham, A. (1979, February). Effect of teacher clarity on student achievement. Paper presented at the meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 177 127)
- Land, M. L., & Smith, L. (1979). The effect of low inference teacher clarity inhibitors on student achievement. Journal of Teacher Education, 31, 55-57.
- Landman, J. T., & Dawes, R. M. (1982). Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. American Psychologist, 37, 504-516.
- Landman, J. T., & Dawes, R. M. (1984). Reply to Orwin and Cordray. American Psychologist, 39, 72-73.
- Leclerc, M., Bertrand, R., & Dufour, N. (1986). Correlations between teaching practices and class achievement in introductory algebra. Teaching and Teacher Education, 2, 355-365.
- Lorentz, J. L. (1977, April). The development of measures of teacher effectiveness from multiple measures of student growth. Paper presented at the meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 137 403)
- Marsh, H. W. (1977). The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal, 14, 441-447.
- Marsh, H. W. (1982). Validity of Students' evaluation of college teaching: A multitrait-multimethod analysis. Journal of Educational Psychology, 74, 264-279.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388.
- Marsh, H. W., Fleiner, H., & Thomas, S. T. (1975). Validity and usefulness of student evaluations of instructional quality. Journal of Educational Psychology, 67, 833-839.

- Marsh, H. W., & Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. Journal of Educational Psychology, 72, 468-475.
- Martikean, A. (1973). The levels of questioning and their effects upon student performance above the knowledge level on Bloom's Taxonomy of Educational Objectives. Gary: Indiana University Northwest (ERIC Document Reproduction Services No. ED 091 248)
- Mathis, P. M., & Shrum, J. W. (1977). The effect of kinetic structure on achievement and total attendance time in audio-tutorial biology. Journal of Research in Science Teaching, 4, 105-115.
- McCaleb, J., & Rosenthal, B. G. (1983). Relationships in teacher clarity between students' perceptions and observers' ratings. Journal of Classroom Interaction, 19(1), 15-21.
- McDonald, F. J., & Elias, P. (1976). The effects of teaching performance on pupil learning (Beginning Teacher Evaluation Study: Phase II 1973-74, Final Report Vol. 1). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Services No. ED 127 364)
- McKeachie, W. J., & Kulik, J. A. (1975). Effective college teaching. In F. N. Kerlinger (Ed.), Review of research in education (Vol: 3, pp. 165-209). Itasca, IL: Peacock.
- McKeachie, W. J., & Linn, Y-G. (1978). A note on validity of student ratings of teaching. Educational Research Quarterly, 4(3), 45-47.
- McKeachie, W. J., Linn, Y-G, & Mann, W. (1971). Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 8, 435-445.
- McKeachie, W. J., Linn, Y-G, & Mendelson, C. A. (1978). A small study assessing teacher effectiveness: Does learning last. Contemporary Educational Psychology, 3, 352-357.
- McKeachie, W. J., & Solomon, D. (1958). Student ratings of instructors: A validity study. Journal of Educational Research, 51, 379-382.
- McKinney, J., Mason, J., Perkerson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. Journal of Educational Psychology, 67, 198-203.

- Medley, D. M., & Mitzel, H. E. (1959). Some behavioral correlates of teacher effectiveness. Journal of Educational Psychology, 49, 86-92.
- Morsh, J. E., Burgess, G. G., & Smith, P. N. (1956). Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 47, 79-88.
- Orpen, C. (1980). Student evaluation of lecturers as an indicator of instructional quality: A validity study. Journal of Educational Research, 74, 5-7.
- Orwin, R. G., & Cordray, D. S. (1984). Smith and Glass's psychotherapy conclusions need further probing: On Landman and Dawes' reanalysis. American Psychologist, 39, 71-72.
- Page, E. B. (1958). Teacher comments and student performance: A seventy-four classroom experiment in school motivation. Journal of Educational Psychology, 49, 173-181.
- Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd Ed.). New York: Holt, Rinehart, & Winston.
- Peterson, D., Micceri, T., & Smith, B. O. (1985). Measurement of teacher performance: A study of instrument development. Teaching and Teacher Education, 1, 63-67.
- Peterson, P. L. (1979). Direct instruction reconsidered. In P. L. Peterson & H. J. Walberg (Eds.), Research on teaching (pp. 57-69). Berkeley, CA: McCutchan.
- Pierce, J. R. (1980). An introduction to information theory: Symbols, signals and noise (2nd ed.). New York: Dover.
- Pinney, R. H. (1970). Presentational behaviors related to success in teaching. Dissertation Abstracts, 30, 5327A. (University Microfilms No. 70-10,552)
- Pitman, R. B. (1985). Perceived instructional effectiveness and associated teaching dimensions. Journal of Experimental Education, 54, 34-39.
- Poonyakanok, P., Thisayakorn, N., & Digby, P. W. (1986). Student evaluation of teacher performance: Some initial research findings from Thailand. Teaching and Teacher Education, 2, 145-154.

- Porter, A. (1985). Do tests and textbooks match? Cap Trends, 11(1), 1-2. (Portland OR: Northwest Regional Educational Laboratory, Center for Performance Assessment)
- Renfrow, D., & Impara, J. C. (1989). Making academic presentations--Effectively! Educational Researcher, 18(2), 20-21.
- Rodin, M., & Rodin, M. (1972). Student evaluation of teachers. Science, 177, 1164-1166.
- Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to measurement of change. Psychological Bulletin, 92, 726-748.
- Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50, 203-228.
- Rosenshine, B. V. (1970a). Enthusiastic teaching: A research review. School Review, 78, 499-514.
- Rosenshine, B. V. (1970b). Evaluation of classroom instruction. Review of Educational Research, 40, 270-301.
- Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. Journal of Educational Psychology, 74, 161-169.
- Ryan, F. L. (1973). Differential effects of levels of questioning on student achievement. Journal of Experimental Education, 41(3), 63-67.
- Ryan, F. L. (1974). The effects on social science achievement of multiple student responding to different levels of questioning. Journal of Experimental Education, 42(4), 71-75.
- Savage, T. V. (1972). A study of the relationship of classroom questions and social studies achievement of fifth-grade children. Dissertation Abstracts International, 33, 2245A. (University Microfilms No. 72-28,661)
- Sharp, C. S. (1966). A study of certain teacher characteristics and behavior as factors affecting pupil achievement in high school biology. Dissertation Abstracts, 27, 1207A-1208A. (University Microfilms No. 66-11,601)

- Shavelson, R., & Dempsey-Atwood, N. (1976). Generalizability of measurement of teacher behavior. Review of Educational Research, 46, 553-612.
- Slavin, R. E. (1984). Meta-analysis in education: How has it been used? Educational Researcher, 13(8), 6-15.
- Smith, L. R. (1977). Aspects of teacher discourse and student achievement in mathematics. Journal of Research in Mathematics Education, 8, 195-204.
- Smith, L. R. (1979). Task oriented lessons and student achievement. Journal of Educational Research, 73, 16-19.
- Smith, L. R. (1985a). Presentational behaviors and student achievement in mathematics. Journal of Educational Research, 78, 292-298.
- Smith, L. R. (1985b). Student perception, teacher clarity, and their relationship to student performance. Educational and Psychological Research, 5, 131-142.
- Smith, L. R. (1985c). Teacher clarifying behaviors: Effects on student achievement and perceptions. Journal of Experimental Education, 53, 162-169.
- Smith, L. R., & Cotton, M. L. (1980). Effect of lesson vagueness and discontinuity on student achievement and attitudes. Journal of Educational Psychology, 72, 670-675.
- Smith, L. R., & Sanders, K. (1981). The effects on student achievement and student perception of varying structure in social studies content. Journal of Educational Research, 74, 333-336.
- Smith, M. L. (1980). Publication bias and meta-analysis. Evaluation in Education, 4, 22-24.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32, 752-760.
- Smith, S. (1978). The identification of teaching behaviors descriptive of the construct: Clarity of presentation. Dissertation Abstracts International, 39, 3529A. (University Microfilms No. 78-23, 593)
- Snider, A. M. (1965). Some relationships between pupil growths in certain basic skills and pupils' perceptions of behaviors of their teachers. Dissertation Abstracts, 26, 3157. (University Microfilms No. 65-11,038)

- Snider, R. M. (1966). A project to study the nature of effective physics teaching. Dissertation Abstracts, 26, 7183. (University Microfilms No. 66-6078)
- Soar, R. S. (1968). Optimum teacher-pupil interaction for pupil growth. Educational Leadership Research Supplement, 26, 275-280.
- Soar, R. S. (1972). Teacher behavior related to pupil growth. International Review of Education, 18(4), 508-526.
- Soar, R. S. (1973). Follow through classroom process measurement and pupil growth (1970-71): Final report. Gainesville, FL: University of Florida, College of Education.
- Soar, R. S., & Soar, R. M. (1973). Classroom behavior, pupil characteristics and pupil growth for the school year and the summer. Gainesville, FL: University of Florida, College of Education.
- Solomon, D., Bezdek, W. E., & Rosenberg, L. (1964). Dimensions of teacher behavior. Journal of Experimental Education, 33, 23-40.
- Solomon, D., & Kendall, A. J. (1976). Individual characteristics and children's performance in varied educational settings. Rockville, MD: Montgomery County Public Schools. (ERIC Document Reproduction Service No. ED 125 958)
- Stallings, J. A. (1974). Follow through classroom observation evaluation 1972-73, Executive summary SRI Project. Washington DC: Office of Education. (ERIC Document Reproduction Service No. ED 104 970)
- Stallings, J. A. (1977). How instructional processes relate to child outcomes. In G. Borich (Ed.), The appraisal of teaching: Concepts and processes (pp. 104-113). Reading, MA: Addison-Wesley.
- Stallings, J. A., & Kaskowitz, D. H. (1974). Follow through classroom observation evaluation. Menlo Park, CA: Stanford Research Institute.
- Sullivan, A. M., & Skanes, G. R. (1974). Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 66, 584-590.
- Tobin, K. G., & Capie, W. (1982). Relationships between classroom process variables and middle-school science achievement. Journal of Educational Psychology, 74, 441-454.

- Torrance, E., & Parent, E. (1966). Characteristics of mathematics teachers that affect students' learning (Cooperative Research Project No. 1020). University of Minnesota, Inst. of Teaching, Minnesota School Math and Science Center. (ERIC Document Reproduction Service No. ED 010 378)
- Trincheró, R. L. (1974). The longitudinal measurement of teacher effectiveness. California Journal of Educational Research, 25, 121-127.
- Trincheró, R. L. (1975). Three technical skills of teaching: Their stability and effect on pupil attitudes and achievement. Dissertation Abstracts International, 36, 5961A.
- Trindade, A. L. (1972). Structures in science teaching and learning outcomes. Journal of Research in Science Teaching, 9, 65-74.
- Turner, R. & Thompson, R. (1974). Relationship between college student ratings of instructors and residual learning. Paper presented at the meeting of the American Educational Research Association, Chicago.
- Vorrayer, D. F. (1969). An analysis of teacher classroom behavior and role. Dissertations Abstracts, 26, 5254. (University Microfilms No. 65-4478)
- Wright, C. J., & Nuthall, G. A. (1970). The relationship between teacher behaviors and pupil achievement in three experimental elementary science lessons. American Educational Research Journal, 7, 477-491.
- Zelby, L. W. (1974). Student-faculty evaluation. Science, 183, 13-17.
- Zimmerman, D. W., & Williams, R. H. (1982). Gain scores in research can be highly reliable. Journal of Educational Measurement, 19, 149-154.

BIOGRAPHICAL SKETCH


I, Frank Fendick, was born in 1930 in married quarters in the "Home of the British Army," Aldershot, England. I attended army schools until I joined the army as a boy soldier in 1945 at the age of 15 (just in time to serve for six months during the war and thus qualify for the War medal). I served in the army for 11 years, mostly as a radio mechanic in the Airborne. I served in Egypt and Cyprus but was discharged, with the rank of sergeant, in February, 1956, so it was not necessary for me to go to jail for refusing to take part in Britain's attack on Egypt in that year.

I attended an agricultural college for a year to obtain my Certificate in Agriculture and worked as a cowman until, with a wife and child to support, I decided that six pounds (perhaps \$18 then) a week did not provide a comfortable living. In 1959 I started at a technical college (equivalent to community college) and in 1961 obtained my A-levels (university entrance qualification) in math and physics. I was then employed at the college to teach these subjects at O-level (two years below A-level) at the princely salary of 700 pounds per year (thus doubling my previous wage).

After three years teaching I entered Queen Mary College, London University, where I obtained a B.Sc. (Honours) in physics. I then taught A-level physics at my previous college for eight years. During this period I obtained, by part-time study, two certificates in education and a Diploma in Education from Leeds University. In partial fulfillment of this qualification I wrote a thesis entitled "The Effects of Teacher-Student Classroom Interaction on Student Achievement and Student Opinion of the Teacher."

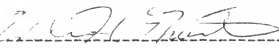
In 1975 my wife left me, so I went to Africa. Instead of joining the Foreign Legion, I taught physics at the University of Maiduguri and at the Federal Advanced Teachers' College, Yola (both in Nigeria), where I was the head of department. By this time I thought I knew quite a lot about teaching and was not impressed by the research that I read on the subject. I, therefore, decided to do some research of my own using the scientific principles that I taught in physics. In 1982 I became a graduate student in the Foundations of Education Department, University of Florida, in order to accomplish this goal. While at the university I have taught half-time in the Physics Department and have been paid \$15,000 a year--more than I have ever earned in my life!

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



James Algina, Chairman
Professor of Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.




Wilson H. Guertin
Professor of Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Patricia T. Ashton
Professor of Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Robert C. Ziller
Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1990

James A. Quinn

Chairman, Foundations of Education

David E. Smith (28)

Dean, College of Education

Dean, Graduate School